Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 5 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Information and Communication Technology Use in Suicide Prevention: Scoping Review (e25288)
Jessica Rassy, Cécile Bardon, Luc Dargis, Louis-Philippe Côté, Laurent Corthésy-Blondin, Carl-Maria Mörch, Réal Labelle. 11

Machine Learning and Natural Language Processing in Mental Health: Systematic Review (e15708)
Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Taylor Ryan, Jonathan Marsh, Jordan DeVylder, Michel Walter, Sofian Berrouiguet, Christophe Lemey. 31

Impact of Asynchronous Electronic Communication–Based Visits on Clinical Outcomes and Health Care Delivery: Systematic Review (e27531)
Oliver Nguyen, Amir Alishahi Tabriz, Jinhai Huo, Karim Hanna, Christopher Shea, Kea Turner. 51

Electronic Tools to Bridge the Language Gap in Health Care for People Who Have Migrated: Systematic Review (e25131)
Frédérique Thonon, Swati Perrot, Abhijna Yergolkar, Olivia Rousset-Torrente, James Griffith, Olivier Chassany, Martin Duracinsky. 65

The Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure, Process, and Outcome (e27180)
Rosian Tossaint-Schoenmakers, Anke Versluis, Niels Chavannes, Esther Talboom-Kamp, Marise Kasteleyn. 79

Social Media Use for Health Purposes: Systematic Review (e17917)
Junhan Chen, Yuan Wang. 102

Maximizing Participant Engagement, Participation, and Retention in Cohort Studies Using Digital Methods: Rapid Review to Inform the Next Generation of Very Large Birth Cohorts (e23499)
Joanna Nkyekyer, Susan Clifford, Fiona Mensah, Yichao Wang, Lauren Chiu, Melissa Wake. 169

Mobile Apps to Improve Medication Adherence in Cardiovascular Disease: Systematic Review and Meta-analysis (e24190)
Shahd Al-Arkee, Julie Mason, Deirdre Lane, Larissa Fabritz, Winnie Chua, M Haque, Zahraa Jalal. 200

Digital Health Strategies for Cervical Cancer Control in Low- and Middle-Income Countries: Systematic Review of Current Implementations and Gaps in Research (e23350)
Andrea Rossman, Hadley Reid, Michelle Pieters, Cecelia Mizelle, Megan von Isenburg, Nimmi Ramanujam, Megan Huchko, Lavanya Vasudevan. 218

The Use of Social Media for Health Research Purposes: Scoping Review (e25736)
Charline Bour, Adrian Ahne, Susanne Schmitz, Camille Perchoux, Coralie Dessenne, Guy Fagherazzi. 234

XSL•FO
RenderX

Theories Informing eHealth Implementation: Systematic Review and Typology Classification (e18500)
Milena Heinsch, Jessica Wyllie, Jamie Carlson, Hannah Wells, Campbell Tickner, Frances Kay-Lambkin. 248

Effectiveness of eHealth Nutritional Interventions for Middle-Aged and Older Adults: Systematic Review and Meta-analysis (e15649)
Caroline Robert, Mojisola Erdt, James Lee, Yuanyuan Cao, Nurhazimah Naharudin, Yin-Leng Theng. 384

Information Quality Frameworks for Digital Health Technologies: Systematic Review (e23479)
Kayode Fadahunsi, Siobhan O'Connor, James Akinlua, Petra Wark, Joseph Gallagher, Christopher Carroll, Josip Car, Azeem Majeed, John O'Donoghue. 855

Agent-Oriented Goal Models in Developing Information Systems Supporting Physical Activity Among Adolescents: Literature Review and Expert Interviews (e24810)
Kerli Mooses, Kuldar Taveter. 878

Priority Given to Technology in Government-Based Mental Health and Addictions Vision and Strategy Documents: Systematic Policy Review (e25547)
Shalini Lal, Lyna Siafa, Hajin Lee, Carol Adair. 955

Original Papers

Effectiveness of Digital Interventions for Reducing Behavioral Risks of Cardiovascular Disease in Nonclinical Adult Populations: Systematic Review of Reviews (e19688)
Natalie Gold, Amy Yau, Benjamin Rigby, Chris Dyke, Elizabeth Remfry, Tim Chadborn. 118

Understanding Public Perceptions of COVID-19 Contact Tracing Apps: Artificial Intelligence–Enabled Social Media Analysis (e26618)
Kathrin Cresswell, Ahsen Tahir, Zakariya Sheikh, Zain Hussain, Andrés Domínguez Hernández, Ewen Harrison, Robin Williams, Aziz Sheikh, Amir Hussain. 277

Initial Outcomes of Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy Tailored to Public Safety Personnel: Longitudinal Observational Study (e27610)
Heather Hadjistavropoulos, Hugh McCall, David Thiessen, Ziyin Huang, R Carleton, Blake Dear, Nickolai Titov. 358

Guided Internet-Based Cognitive Behavioral Therapy for Depression: Implementation Cost-Effectiveness Study (e27410)
Jordi Piera-Jiménez, Anne Etzelmueller, Spyros Kolovos, Frans Folkvord, Francisco Lupiáñez-Villanueva. 371

Internet-Based Cognitive Behavioral Therapy for Patients Reporting Symptoms of Anxiety and Depression After Myocardial Infarction: U-CARE Heart Randomized Controlled Trial Twelve-Month Follow-up (e25465)
Sophia Humphries, John Wallert, Fredrika Norlund, Emma Wallin, Gunilla Burell, Louise von Essen, Claes Held, Erik Olsson. 400

Intelligent Personalized Exercise Prescription Based on an eHealth Promotion System to Improve Health Outcomes of Middle-Aged and Older Adult Community Dwellers: Pretest–Posttest Study (e28221)
Ting Sun, Yang Xu, Hui Xie, Zuchang Ma, Yu Wang. 414

A Social Media Organizational Productivity Model: Insights From Public Health Professionals (e23792)
Hamad Dailah Sr, Muhammad Naeem. 428

Sleep Disturbances in Frontline Health Care Workers During the COVID-19 Pandemic: Social Media Survey Study (e27331)
Nancy Stewart, Anya Koza, Serena Dhaon, Christiana Shoushtari, Maylyn Martinez, Vineet Arora. 441

XSL•FO
RenderX

Developing an Automatic System for Classifying Chatter About Health Services on Twitter: Case Study for Medicaid (e26616)
Yuan-Chi Yang, Mohammed Al-Garadi, Whitney Bremer, Jane Zhu, David Grande, Abeed Sarker. 451

Information-Seeking Patterns During the COVID-19 Pandemic Across the United States: Longitudinal Analysis of Google Trends Data (e22933)
Tichakunda Mangono, Peter Smittenaar, Yael Caplan, Vincent Huang, Staci Sutermaster, Hannah Kemp, Sema Sgaier. 464

Mining and Validating Social Media Data for COVID-19–Related Human Behaviors Between January and July 2020: Infodemiology Study (e27059)
Ashlynn Daughton, Courtney Shelley, Martha Barnard, Dax Gerts, Chrysm Watson Ross, Isabel Crooker, Gopal Nadiga, Nilesh Mukundan, Nidia Vaquera Chavez, Nidhi Parikh, Travis Pitts, Geoffrey Fairchild. 480

Detecting Impending Stroke From Cognitive Traits Evident in Internet Searches: Analysis of Archival Data (e27084)
Sigal Shaklai, Ran Gilad-Bachrach, Elad Yom-Tov, Naftali Stern. 494

The Uptake and Use of Telemonitoring in Chronic Care Between 2014 and 2019: Nationwide Survey Among Patients and Health Care Professionals in the Netherlands (e24908)
Martine Huygens, Helene Voogdt-Pruis, Myrah Wouters, Maaike Meurs, Britt van Lettow, Conchita Kleijweg, Roland Friele. 504

The Costs and Cardiovascular Benefits in Patients With Peripheral Artery Disease From a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study (e24346)
Jen-Kuang Lee, Chi-Sheng Hung, Ching-Chang Huang, Ying-Hsien Chen, Hui-Wen Wu, Pao-Yu Chuang, Jiun-Yu Yu, Yi-Lwun Ho. 515

Characteristics of Citizens and Their Use of Teleconsultations in Primary Care in the Catalan Public Health System Before and During the COVID-19 Pandemic: Retrospective Descriptive Cross-sectional Study (e28629)
Oscar Solans, Josep Vidal-Alaball, Pasqual Roig Cabo, Núria Mora, Ermengol Coma, Josep Bonet Simó, Eduardo Hermosilla Pérez, Francesc Saigí-Rubió, Carmen Olmos Domínguez, Jordi Piera-Jiménez, Mercè Abizanda González, Francesc López Seguí. 533

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 (e27918)
Patricia Areán, Abhishek Pratap, Honor Hsin, Tierney Huppert, Karin Hendricks, Patrick Heagerty, Trevor Cohen, Courtney Bagge, Katherine Comtois. 544

Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions (e26494)
Akira-Sebastian Poncette, Maximilian Wunderlich, Claudia Spies, Patrick Heeren, Gerald Vorderwülbecke, Eduardo Salgado, Marc Kastrup, Markus Feufel, Felix Balzer. 557

Integrating Option Grid Patient Decision Aids in the Epic Electronic Health Record: Case Study at 5 Health Systems (e22766)
Peter Scalia, Farhan Ahmad, Danielle Schubbe, Rachel Forcino, Marie-Anne Durand, Paul Barr, Glyn Elwyn. 569

Patient Portal Use, Perceptions of Electronic Health Record Value, and Self-Rated Primary Care Quality Among Older Adults: Cross-sectional Survey (e22549)
Dori Cross, Zachary Levin, Minakshi Raj. 583

Evaluating the Effectiveness of NoteAid in a Community Hospital Setting: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Patients (e26354)
John Lalor, Wen Hu, Matthew Tran, Hao Wu, Kathleen Mazor, Hong Yu. 596

Patient Rationales Against the Use of Patient-Accessible Electronic Health Records: Qualitative Study (e24090)
Hanne Valeur, Anne Lie, Kåre Moen. 608

XSL•FO
RenderX

Testing a Novel Web-Based Neurocognitive Battery in the General Community: Validation and Usability Study (e25082)
Riley Capizzi, Melissa Fisher, Bruno Biagianti, Neelufaer Ghiasi, Ariel Currie, Karrie Fitzpatrick, Nicholas Albertini, Sophia Vinogradov. 618

Effects of an Animated Blood Clot Technology (Visual Clot) on the Decision-Making of Users Inexperienced in Viscoelastic Testing: Multicenter Trial (e27124)
Sadiq Said, Tadzio Roche, Julia Braun, Micheal Ganter, Patrick Meybohm, Johannes Herrmann, Kai Zacharowski, Florian Raimann, Florian Piekarski, Eva Rivas, Manuel López-Baamonde, Donat Spahn, Christoph Nöthiger, David Tscholl. 631

WeChat as a Platform for Problem-Based Learning Among Hematological Postgraduates: Feasibility and Acceptability Study (e16463)
Ping Luo, Wenwen Pang, Yingying Wang, Minghui Liu, Shu Zhou, Shuai Liu, Xian Zhang, Li Liu, Yanan Liu, Fuling Zhou. 643

Influences of Medical Crowdfunding Website Design Features on Trust and Intention to Donate: Controlled Laboratory Experiment (e25554)
Xing Zhang, Wenli Hu, Quan Xiao. 660

Willingness to Adopt Health Information Among Social Question-and-Answer Community Users in China: Cross-sectional Survey Study (e27811)
PengFei Li, Lin Xu, Tingting Tang, Xiaoqian Wu, Cheng Huang. 675

A Clinical Prediction Model to Predict Heparin Treatment Outcomes and Provide Dosage Recommendations: Development and Validation Study (e27118)
Dongkai Li, Jianwei Gao, Na Hong, Hao Wang, Longxiang Su, Chun Liu, Jie He, Huizhen Jiang, Qiang Wang, Yun Long, Weiguo Zhu. 688

Automatic Classification of Screen Gaze and Dialogue in Doctor-Patient-Computer Interactions: Computational Ethnography Algorithm Development and Validation (e25218)
Samar Helou, Victoria Abou-Khalil, Riccardo Iacobucci, Elie El Helou, Ken Kiyono. 701

Usability of Rapid Cholera Detection Device (OmniVis) for Water Quality Workers in Bangladesh: Iterative Convergent Mixed Methods Study (e22973)
Theresa Rager, Cristian Koepfli, Wasif Khan, Sabeena Ahmed, Zahid Mahmud, Katherine Clayton. 715

Patient Recruitment Into a Multicenter Clinical Cohort Linking Electronic Health Records From 5 Health Systems: Cross-sectional Analysis (e24003)
Wendy Bennett, Carolyn Bramante, Scott Rothenberger, Jennifer Kraschnewski, Sharon Herring, Michelle Lent, Jeanne Clark, Molly Conroy, Harold Lehmann, Nickie Cappella, Megan Gauvey-Kern, Jody McCullough, Kathleen McTigue. 729

Emotional Reactions to Cybersecurity Breach Situations: Scenario-Based Survey Study (e24879)
Sanja Budimir, Johnny Fontaine, Nicole Huijts, Antal Haans, George Loukas, Etienne Roesch. 740

Understanding Barriers to Novel Data Linkages: Topic Modeling of the Results of the LifeInfo Survey (e24236)
Holly Clarke, Stephen Clark, Mark Birkin, Heather Iles-Smith, Adam Glaser, Michelle Morris. 754

HIV Information Acquisition and Use Among Young Black Men Who Have Sex With Men Who Use the Internet: Mixed Methods Study (e22986)
Megan Threats, Keosha Bond. 773

Validation of 4D Components for Measuring Quality of the Public Health Data Collection Process: Elicitation Study (e17240)
Hong Chen, Ping Yu, David Hailey, Tingru Cui. 790

XSL•FO
RenderX

Feasibility, Acceptability, and Efficacy of Virtual Reality Training for Older Adults and People With Disabilities: Single-Arm Pre-Post Study (e27640)
Pui Chau, Yan Kwok, Mee Chan, Ka Kwan, Kam Wong, Ying Tang, Kan Chau, Sheung Lau, Yan Yiu, Mei Kwong, Wai Lai, Mun Leung. 806

Effect of Cognitive Training in Fully Immersive Virtual Reality on Visuospatial Function and Frontal-Occipital Functional Connectivity in Predementia: Randomized Controlled Trial (e24526)
Jae Kang, Nambeom Kim, Sook Lee, Soo Woo, Geumjin Park, Byeong Yeon, Jung Park, Jung-Hae Youn, Seung-Ho Ryu, Jun-Young Lee, Seong-Jin Cho. 822

Assessing the Feasibility of an Open-Source Virtual Reality Mirror Visual Feedback Module for Complex Regional Pain Syndrome: Pilot Usability Study (e16536)
Andrea Won, Ariana Barreau, Mark Gaertner, Tristan Stone, Joshua Zhu, Cheng Wang, Sean Mackey. 840

Virtual Reality for Sedation During Atrial Fibrillation Ablation in Clinical Practice: Observational Study (e26349)
Thomas Roxburgh, Anthony Li, Charles Guenancia, Patrice Pernollet, Claire Bouleti, Benjamin Alos, Matthieu Gras, Thomas Kerforne, Denis Frasca, François Le Gal, Luc Christiaens, Bruno Degand, Rodrigue Garcia. 848

Digitization of Measurement-Based Care Pathways in Mental Health Through REDCap and Electronic Health Record Integration: Development and Usability Study (e25656)
Steve Hawley, Joanna Yu, Nikola Bogetic, Natalia Potapova, Chris Wakefield, Mike Thompson, Stefan Kloiber, Sean Hill, Damian Jankowicz, David Rotenberg. 867

Guided Internet-Based Cognitive Behavioral Therapy for Insomnia: Health-Economic Evaluation From the Societal and Public Health Care Perspective Alongside a Randomized Controlled Trial (e25609)
Claudia Buntrock, Dirk Lehr, Filip Smit, Hanne Horvath, Matthias Berking, Kai Spiegelhalder, Heleen Riper, David Ebert. 895

Medical Specialty Recommendations by an Artificial Intelligence Chatbot on a Smartphone: Development and Deployment (e27460)
Hyeonhoon Lee, Jaehyun Kang, Jonghyeon Yeo. 910

Addressing Biodisaster X Threats With Artificial Intelligence and 6G Technologies: Literature Review and Critical Insights (e26109)
Zhaohui Su, Dean McDonnell, Barry Bentley, Jiguang He, Feng Shi, Ali Cheshmehzangi, Junaid Ahmad, Peng Jia. 920

Personalization of Conversational Agent-Patient Interaction Styles for Chronic Disease Management: Two Consecutive Cross-sectional Questionnaire Studies (e26643)
Christoph Gross, Theresa Schachner, Andrea Hasl, Dario Kohlbrenner, Christian Clarenbach, Forian Wangenheim, Tobias Kowatsch. 936

The Patterns and Impact of Social Media Exposure of Journal Publications in Gastroenterology: Retrospective Cohort Study (e25252)
Austin Chiang, Loren Rabinowitz, Javid Alakbarli, Walter Chan. 969

Evaluating Scholars' Impact and Influence: Cross-sectional Study of the Correlation Between a Novel Social Media–Based Score and an Author-Level Citation Metric (e28859)
Lucas Oliveira J e Silva, Graciela Maldonado, Tara Brigham, Aidan Mullan, Audun Utengen, Daniel Cabrera. 977

Leveraging Genomic Associations in Precision Digital Care for Weight Loss: Cohort Study (e25401)
Ranjan Sinha, Dashyanng Kachru, Roshni Ricchetti, Simitha Singh-Rambiritch, Karthik Muthukumar, Vidhya Singaravel, Carmel Irudayanathan, Chandana Reddy-Sinha, Imran Junaid, Garima Sharma, Patricia Francis-Lyon. 987

Analysis of COVID-19 Transmission Sources in France by Self-Assessment Before and After the Partial Lockdown: Observational Study (e26932)
Fabrice Denis, Anne-Lise Septans, Florian Le Goff, Stephan Jeanneau, François-Xavier Lescure. 1038

XSL•FO
RenderX

Using a Secure, Continually Updating, Web Source Processing Pipeline to Support the Real-Time Data Synthesis and Analysis of Scientific Literature: Development and Validation Study (e25714)
Uddhav Vaghela, Simon Rabinowicz, Paris Bratsos, Guy Martin, Epameinondas Fritzilas, Sheraz Markar, Sanjay Purkayastha, Karl Stringer, Harshdeep Singh, Charlie Llewellyn, Debabrata Dutta, Jonathan Clarke, Matthew Howard, PanSurg REDASA Curators, Ovidiu Serban, James Kinross. . . . 1045

Associations Between Health Literacy, eHealth Literacy, and COVID-19–Related Health Behaviors Among Chinese College Students: Cross-sectional Online Study (e25600)
Shaojie Li, Guanghui Cui, Atipatsa Kaminga, Sixiang Cheng, Huilan Xu. . . . 1059

Understanding Behavioral Intentions Toward COVID-19 Vaccines: Theory-Based Content Analysis of Tweets (e28118)
Siru Liu, Jialin Liu. . . . 1072

Influence of COVID-19 Lockdowns on the Usage of a Vision Assistance App Among Global Users With Visual Impairment: Big Data Analytics Study (e26283)
Gang Luo, Shrinivas Pundlik. . . . 1082

The Efficacy of Computerized Cognitive Behavioral Therapy for Depressive and Anxiety Symptoms in Patients With COVID-19: Randomized Controlled Trial (e26883)
Zhifen Liu, Dan Qiao, Yifan Xu, Wentao Zhao, Yang Yang, Dan Wen, Xinrong Li, Xiaoping Nie, Yongkang Dong, Shiyou Tang, Yi Jiang, Ying Wang, Juan Zhao, Yong Xu. . . . 1092

Factors Influencing Anxiety Among WeChat Users During the Early Stages of the COVID-19 Pandemic in Mainland China: Cross-sectional Survey Study (e24412)
Changqing Zou, Weiyu Zhang, Kristin Sznajder, Fengzhi Yang, Yajing Jia, Ruqing Ma, Can Cui, Xiaoshi Yang. . . . 1107

Experiences and Attitudes of Elementary School Students and Their Parents Toward Online Learning in China During the COVID-19 Pandemic: Questionnaire Study (e24496)
Shu Cui, Chao Zhang, Shijiang Wang, Xingong Zhang, Lei Wang, Ling Zhang, Qiuyu Yuan, Cui Huang, Fangshuo Cheng, Kai Zhang, Xiaoqin Zhou. . . . 1125

Utilizing Health Behavior Change and Technology Acceptance Models to Predict the Adoption of COVID-19 Contact Tracing Apps: Cross-sectional Survey Study (e25447)
Samuel Tomczyk, Simon Barth, Silke Schmidt, Holger Muehlan. . . . 1137

Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis (e26953)
Stephen Kwok, Sai Vadde, Guanjin Wang. . . . 1156

Characterizing Health Care Delays and Interruptions in the United States During the COVID-19 Pandemic: Internet-Based, Cross-sectional Survey Study (e25446)
Elizabeth Papautsky, Dylan Rice, Hana Ghoneima, Anna McKowen, Nicholas Anderson, Angie Wootton, Cindy Veldhuis. . . . 1172

Capturing COVID-19–Like Symptoms at Scale Using Banner Ads on an Online News Platform: Pilot Survey Study (e24742)
Brian Dixon, Sumit Mukherjee, Ashley Wiensch, Mary Gray, Juan Ferres, Shaun Grannis. . . . 1188

An App-Based Intervention to Support First Responders and Essential Workers During the COVID-19 Pandemic: Needs Assessment and Mixed Methods Implementation Study (e26573)
Stacie Vilendrer, Alexis Amano, Cati Brown Johnson, Marissa Favet, Nadia Safaeinili, Jacqueline Villasenor, Jonathan Shaw, Attila Hertelendy, Steven Asch, Megan Mahoney. . . . 1202

XSL•FO
RenderX

A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study (e27806)
Cheng-Sheng Yu, Shy-Shin Chang, Tzu-Hao Chang, Jenny Wu, Yu-Jiun Lin, Hsiung-Fei Chien, Ray-Jade Chen. 1219

Characterization of an Open-Access Medical News Platform's Readership During the COVID-19 Pandemic: Retrospective Observational Study (e26666)
Alex Chan, Constance Wu, Andrew Cheung, Marc Succi. 1238

Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups Among Patients with COVID-19: Observational Longitudinal Study (e25988)
Julián Benito-León, Mª del Castillo, Alberto Estirado, Ritwik Ghosh, Souvik Dubey, J Serrano. 1245

Olfactory Training and Visual Stimulation Assisted by a Web Application for Patients With Persistent Olfactory Dysfunction After SARS-CoV-2 Infection: Observational Study (e29583)
Fabrice Denis, Anne-Lise Septans, Lea Periers, Jean-Michel Maillard, Florian Legoff, Hirac Gurden, Sylvain Moriniere. 1259

A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score (e29058)
Geza Halasz, Michela Sperti, Matteo Villani, Umberto Michelucci, Piergiuseppe Agostoni, Andrea Biagi, Luca Rossi, Andrea Botti, Chiara Mari, Marco Maccarini, Filippo Pura, Loris Roveda, Alessia Nardecchia, Emanuele Mottola, Massimo Nolli, Elisabetta Salvioni, Massimo Mapelli, Marco Deriu, Dario Piga, Massimo Piepoli. 1268

Age-Stratified Infection Probabilities Combined With a Quarantine-Modified Model for COVID-19 Needs Assessments: Model Development Study (e19544)
Vena Bongolan, Jose Minoza, Romulo de Castro, Jesus Sevilleja. 1281

Perceptions of Mobile Health Apps and Features to Support Psychosocial Well-being Among Frontline Health Care Workers Involved in the COVID-19 Pandemic Response: Qualitative Study (e26282)
Sungwon Yoon, Hendra Goh, Gayathri Nadarajan, Sharon Sung, Irene Teo, Jungup Lee, Marcus Ong, Nicholas Graves, Tess Teo. 1287

Pain Recognition With Electrocardiographic Features in Postoperative Patients: Method Validation Study (e25079)
Emad Kasaeyan Naeini, Ajan Subramanian, Michael-David Calderon, Kai Zheng, Nikil Dutt, Pasi Liljeberg, Sanna Salantera, Ariana Nelson, Amir Rahmani. 1300

Viewpoints

Use of Twitter in Neurology: Boon or Bane? (e25229)
Biswamohan Mishra, Monica Saini, Carolynne Doherty, Robert Pitceathly, Roopa Rajan, Omar Siddiqi, Gita Ramdharry, Ajay Asranna, Pedro Tomaselli, Allan Kermode, Jawad Bajwa, Divyani Garg, Venugopalan Vishnu. 187

COVID-19, Social Isolation, and Mental Health Among Older Adults: A Digital Catch-22 (e21864)
Theodore Cosco, Karen Fortuna, Andrew Wister, Indira Riadi, Kevin Wagner, Andrew Sixsmith. 265

Enhancing Data Linkage to Break the Chain of COVID-19 Spread: The Taiwan Experience (e24294)
Wei-Cheng Lo, Fu-Chung Wang, Li-Yin Lin, Hong-Wei Jyan, Hsuan-Chien Wu, Yao-Liang Huang, I-Ming Parng, Hung-Yi Chiou. 268

An eHealth Framework for Managing Pediatric Growth Disorders and Growth Hormone Therapy (e27446)
Paul Dimitri, Luis Fernandez-Luque, Indraneel Banerjee, Ignacio Bergadá, Luis Calliari, Jovanna Dahlgren, Antonio de Arriba, Risto Lapatto, Thomas Reinehr, Senthil Senniappan, Cécile Thomas-Teinturier, Meng-Che Tsai, Azriyanti Anuar Zaini, Merat Bagha, Ekaterina Koledova. . .
2 8 5

Bots and Misinformation Spread on Social Media: Implications for COVID-19 (e26933)
McKenzie Himelein-Wachowiak, Salvatore Giorgi, Amanda Devoto, Muhammad Rahman, Lyle Ungar, H Schwartz, David Epstein, Lorenzo Leggio, Brenda Curtis. 298

XSL•FO
RenderX

A Real-World Rheumatology Registry and Research Consortium: The German RheumaDatenRhePort (RHADAR) Registry (e28164)
Stefan Kleinert, Peter Bartz-Bazzanella, Cay von der Decken, Johannes Knitza, Torsten Witte, Sándor Fekete, Matthias Konitzny, Alexander Zink, Georg Gauler, Patrick Wurth, Peer Aries, Kirsten Karberg, Christoph Kuhn, Florian Schuch, Susanna Späthling-Mestekemper, Wolfgang Vorbrüggen, Matthias Englbrecht, Martin Welcker, RHADAR Group. . . . 309

Artificial Intelligence Can Improve Patient Management at the Time of a Pandemic: The Role of Voice Technology (e22959)
Tomasz Jadczyk, Wojciech Wojakowski, Michal Tendera, Timothy Henry, Gregory Egnaczyk, Satya Shreenivas. . . . 318

Improving the Usability and Safety of Digital Health Systems: The Role of Predictive Human-Computer Interaction Modeling (e25281)
Chris Paton, Andre Kushniruk, Elizabeth Borycki, Mike English, Jim Warren. . . . 329

The Impact of Inpatient Telemedicine on Personal Protective Equipment Savings During the COVID-19 Pandemic: Cross-sectional Study (e28845)
Reem Halabi, Geoffrey Smith, Marc Sylwestrzak, Brian Clay, Christopher Longhurst, Lina Lander. . . . 528

Transitioning to Web-Based Learning in Basic Life Support Training During the COVID-19 Pandemic to Battle the Fear of Out-of-Hospital Cardiac Arrest: Presentation of Novel Methods (e27108)
Katarzyna Naylor, Kamil Torres. . . . 650

SARS-CoV-2: The Second Wave in Europe (e22431)
Athanassios Fokas, George Kastis. . . . 1120

Tutorials

Determination of Patient Sentiment and Emotion in Ophthalmology: Infoveillance Tutorial on Web-Based Health Forum Discussions (e20803)
Anne Nguyen, Xuan-Vi Trinh, Sophia Wang, Albert Wu. . . . 337

Establishing and Facilitating Large-Scale Manuscript Collaborations via Social Media: Novel Method and Tools for Replication (e25077)
Kimberly Acquaviva. . . . 350

Letters to the Editor

COVID-19 as a “Force Majeure” for Non–COVID-19 Clinical and Translational Research. Comment on “Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study” (e27937)
Petar Milovanovic, Igor Dumic. . . . 1007

Authors' Reply to: COVID-19 as a “Force Majeure” for Non–COVID-19 Clinical and Translational Research. Comment on “Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study” (e29156)
Andreas Älgå, Oskar Eriksson, Martin Nordberg. . . . 1010

Methodological Clarifications and Generalizing From Weibo Data. Comment on “Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo” (e26255)
Om Yadav. . . . 1012

XSL•FO
RenderX

Authors' Reply to: Methodological Clarifications and Generalizing From Weibo Data. Comment on "Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo" (e29145)
Zhuo-Ying Tao, Yu-Xiong Su. . . . 1014

Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study" (e26148)
Elizabeth Millen, Andreas Gilsdorf, Matthew Fenech, Stephen Gilbert. . . . 1017

Authors' Reply to: Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study" (e26543)
Nicolas Munsch, Alistair Martin, Stefanie Gruarin, Jama Nateqi, Isselmou Abdarahmane, Rafael Weingartner-Ortner, Bernhard Knapp. . . . 1019

Minimizing Selection and Classification Biases. Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing" (e27142)
Francisco Martos Pérez, Ricardo Gomez Huelgas, María Martín Escalante, José Casas Rojo. . . . 1021

Authors' Reply to: Minimizing Selection and Classification Biases Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing" (e29405)
Jose Izquierdo, Joan Soriano. . . . 1023

Objective Outcomes Evaluation of Innovative Digital Health Curricula. Comment on "Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study" (e26034)
Alexander Grzeska, Shan Ali, Tomasz Szmuda, Paweł Słoniewski. . . . 1025

Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings" (e28666)
Daniel Campos, Umberto Fulco, Jonas Oliveira. . . . 1027

Authors' Reply to: Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings" (e29507)
Lauge Rasmussen, Ole Norgaard, Tue Andersen, Adam Palayew, Joey Nicholson, Jeffrey Lazarus. . . . 1030

Corrigenda and Addendas

Correction: Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study (e30101)
Stanley Xu, Sungching Glenn, Lina Sy, Lei Qian, Vennis Hong, Denison Ryan, Steven Jacobsen. . . . 1032

Correction: Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study (e30215)
Malte Schmieding, Rudolf Mörgeli, Maike Schmieding, Markus Feufel, Felix Balzer. . . . 1034

Correction: Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey (e30100)
Fengyun Yu, Pascal Geldsetzer, Anne Meierkord, Juntao Yang, Qiushi Chen, Lirui Jiao, Nadeem Abou-Arraj, An Pan, Chen Wang, Till Bärnighausen, Simiao Chen. . . . 1036

XSL•FO
RenderX

Short Paper

Telehealth Use by Age and Race at a Single Academic Medical Center During the COVID-19 Pandemic: Retrospective Cohort Study (e23905)
Jennifer Stevens, Oren Mechanic, Lawrence Markson, Ashley O'Donoghue, Alexa Kimball. 1184

XSL•FO
RenderX

Review

Information and Communication Technology Use in Suicide Prevention: Scoping Review

Jessica Rassy[1,2,3,4*], PhD; Cécile Bardon[1,5*], PhD; Luc Dargis[1*], MSc; Louis-Philippe Côté[1,5*], MSc; Laurent Corthésy-Blondin[1,5*], MSc; Carl-Maria Mörch[1,5,6,7*], PhD; Réal Labelle[1,2,5,8], PhD

[1]Center for Research and Intervention on Suicide, Ethical Issues and End-of-Life Practices, Université du Québec à Montréal, Montréal, QC, Canada
[2]Research Center, Institut universitaire en santé mentale de Montréal, Montréal, QC, Canada
[3]School of Nursing, Université de Sherbrooke, Longueuil, QC, Canada
[4]Quebec Network on Nursing Intervention Research, Montréal, QC, Canada
[5]Department of Psychology, Université du Québec à Montréal, Montréal, QC, Canada
[6]Algora Lab, Université de Montréal, Montréal, QC, Canada
[7]Mila, Quebec Artificial Intelligence Institute, Montréal, QC, Canada
[8]Department of Psychiatry, Université de Montréal, Montréal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
Réal Labelle, PhD
Department of Psychology
Université du Québec à Montréal
CP 8888, succ Centre-Ville
Montréal, QC, H3C 3P8
Canada
Phone: 1 514 987 4184
Email: labelle.real@uqam.ca

Abstract

Background: The use of information and communication technology (ICT) in suicide prevention has progressed rapidly over the past decade. ICT plays a major role in suicide prevention, but research on best and promising practices has been slow.

Objective: This paper aims to explore the existing literature on ICT use in suicide prevention to answer the following question: what are the best and most promising ICT practices for suicide prevention?

Methods: A scoping search was conducted using the following databases: PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore. These databases were searched for articles published between January 1, 2013, and December 31, 2018. The five stages of the scoping review process were as follows: identifying research questions; targeting relevant studies; selecting studies; charting data; and collating, summarizing, and reporting the results. The World Health Organization suicide prevention model was used according to the continuum of universal, selective, and indicated prevention.

Results: Of the 3848 studies identified, 115 (2.99%) were selected. Of these, 10 regarded the use of ICT in universal suicide prevention, 53 referred to the use of ICT in selective suicide prevention, and 52 dealt with the use of ICT in indicated suicide prevention.

Conclusions: The use of ICT plays a major role in suicide prevention, and many promising programs were identified through this scoping review. However, large-scale evaluation studies are needed to further examine the effectiveness of these programs and strategies. In addition, safety and ethics protocols for ICT-based interventions are recommended.

(J Med Internet Res 2021;23(5):e25288) doi:10.2196/25288

KEYWORDS

suicide prevention; information and communication technology; scoping review; mobile phone

XSL•FO
RenderX

Introduction

Background

Information and communication technology (ICT) has been used for suicide prevention over the past decade. Moreover, there is a growing body of evidence supporting the use of ICT in the development of promising suicide prevention practices [1,2]. ICT can be used to screen individuals at risk of suicide on the web; offer information and help regarding suicidal thoughts and behavior; and offer web-based assessment, interventions, and follow-up [1-6]. The use of ICT widens accessibility to hard-to-reach individuals who do not always seek help in person and offers treatment opportunities to communities with lower access to care, such as rural communities. The use of ICT in suicide prevention can also help professionals offer better care to their patients by combining multiple approaches, such as using mobile apps to monitor symptoms or providing web-based therapeutic programs [7,8].

Many opportunities arise when using ICT to expand suicide prevention strategies. However, the pace at which ICT is advancing makes it difficult to keep up with, especially from a research perspective. To make better use of these different web-based suicide prevention strategies, a better understanding of these different uses of ICT is necessary. We identified 3 major questions. First, a better understanding of how to emphasize the technical aspects of ICT regarding the development, maintenance, privacy, and life cycle of the technology would help make better choices regarding which ICT to use in what context. For example, machine learning offers many possibilities for identifying at-risk individuals, but major technical and ethical considerations must be described and taken into account. Second, we need to improve our understanding of ICT use in suicide prevention. To support adequate decision making, it is important to know whether ICT-based suicide prevention strategies reach different clienteles or if we are reaching the same individuals differently. Third, a better understanding of the structure and efficacy of various types of current web-based assessments and interventions is also important. For instance, can artificial intelligence (AI) and machine learning be used to optimize and accelerate the assessment of individuals at risk? Is a suicide crisis intervention by chat as effective as talking over the phone or in person? Are web-based cognitive behavioral therapy (CBT) programs as effective for suicide prevention as face-to-face CBT programs? Before shifting toward using and recommending the use of these different ICT-based assessments and interventions, it should be noted that their effectiveness has yet to be demonstrated in the literature.

Current Study

The available literature often refers to a specific type of ICT as opposed to a general overview of existing ICT-based strategies in mental health and suicide prevention. More precisely, existing literature reviews have focused on specific ICTs such as smartphone tools [1], web-based interventions [3], mobile apps [4], and social media [5] or specific age groups such as youth [6]. Based on this background and a rapidly evolving corpus of research, we set out to review the evidence supporting the use of all types of ICTs in different levels of suicide prevention strategies. More precisely, we carried out a scoping review to identify the best and most promising practices for ICT use at all levels of suicide prevention, which are described by the World Health Organization as the universal (entire population), selective (specific subpopulations), and indicated (high-risk individuals) levels [9]. ICT was defined as all materials, software, or services used to collect, process, and transmit information; this includes electronic, computer, telecommunications, multimedia, and internet technologies [10]. AI was defined as all theories and techniques used to develop machines capable of simulating intelligence [7].

Methods

Scoping Review Framework

We used the 5-stage scoping framework developed by Arksey and O'Malley [11] and adapted by Levac et al [12] to review peer-reviewed publications. These stages are (1) identifying research questions; (2) identifying relevant studies; (3) selecting studies; (4) charting data; and (5) collating, summarizing, and reporting results. A scoping review allows researchers to explore the available data on various aspects of a theme instead of just one [11,12]. This provides an overview of the relevant literature that helps to identify the various ICT and suicide prevention strategies examined by researchers since 2013 and to document their effects. The Center for Research and Intervention on Suicide, Ethical Issues, and End-of-Life Practices decided to better understand the use of ICT in suicide prevention as of 2013 following the publication of the book *Suicide prevention and new technologies: Evidence-based practice* [13]. This book provides an overview of new technologies in suicide prevention and demonstrates the need for further research in this area. As ICT tools, contents, and use evolve rapidly, a 5-year timeframe seems relevant to address the current state of ICT preventive practices and research. For example, major social networks have changed their policies toward content related to suicidality. During the 2019 International Suicide Prevention Day, Facebook revised its policy of preventing self-harm and suicide by banning graphic representations of suicidality, which could change the level of exposure of individuals vulnerable to suicide. At the same time, the rapid development of scientific knowledge and technology has been noted many times. For example, a review of the scientific literature published in 2013 identified 5 scientifically evaluated mobile apps for child and adolescent mental health [14], whereas Grist et al [15] identified 19 more. In addition, web-based platforms such as Parler or 4Chan were not as widely known and used 10 years ago. Including older technologies and neglecting to consider the rapid content change of internet content over the years may skew results toward obsolete prevention strategies.

Summary of Search Strategy

Identifying the Research Question

The research question identified for this scoping review aims to guide future program development and implementation by establishing an inventory of strategies using ICT that have been the subject of research in recent years. Following the PRISMA

XSL•FO
RenderX

(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations, the research question was developed using the PICO (population, intervention, comparison, and outcomes) conceptual tool [16]. Note that an adaptation of PICO was made through an amalgamation with the SPIDER (sample, phenomenon of interest, design, evaluation, research type) tool by Methley et al [17] to facilitate the consideration of the psychosocial nature of the research object. Consequently, the research question was "How information and communication technology (ICT) can be used in intervention and prevention of suicide among the general population, people at risk, or suicidal people of all ages and conditions combined to contribute to the reduction of any indicators relating to suicidality?" This question was selected to promote the instrumental use of the research results [18].

Identifying Relevant Studies

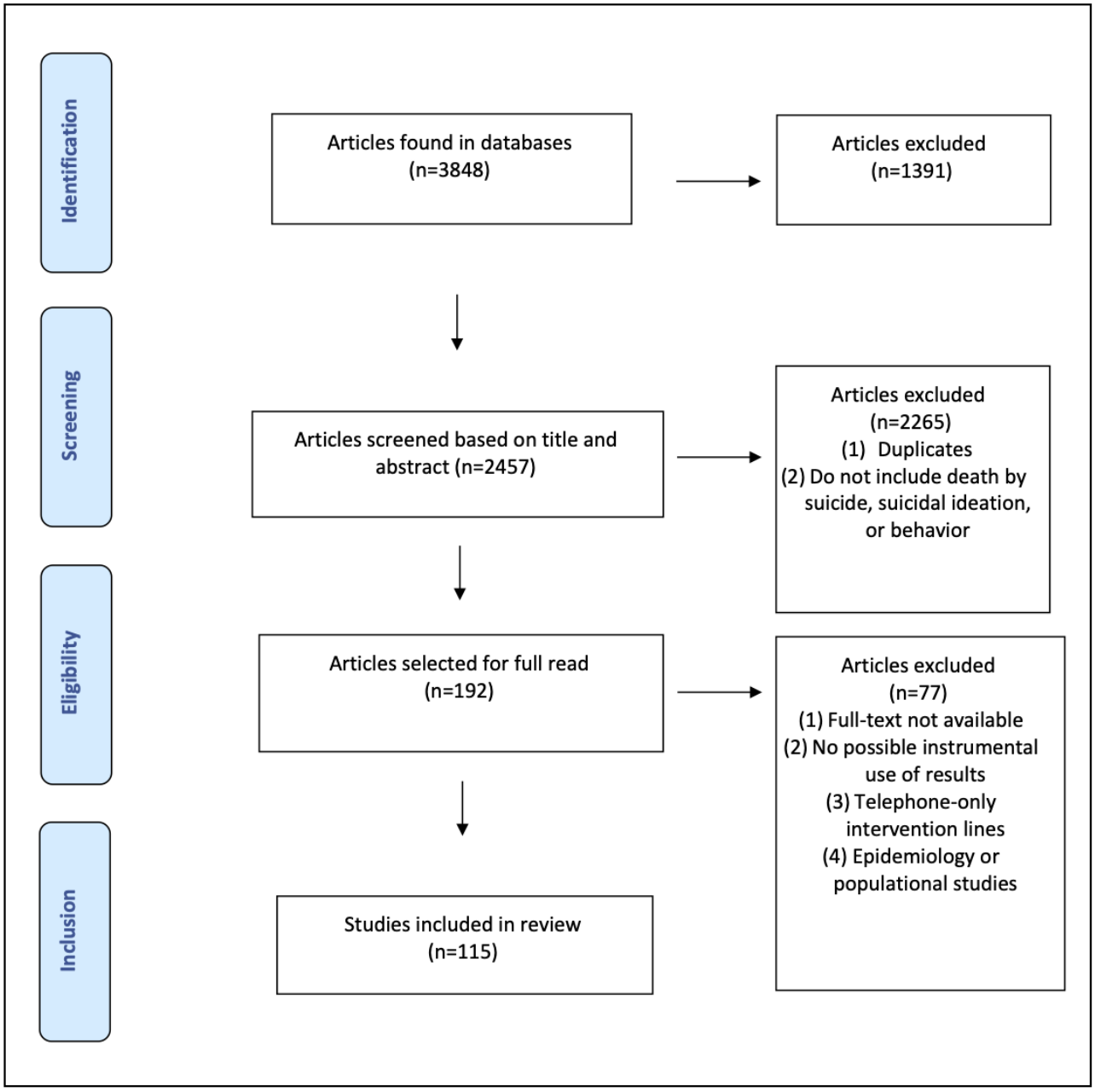
The search strategy consisted of identifying key concepts related to ICT-based suicide prevention and intervention, developing a provisional syntax specific to each database for pretesting, and determining the definitive syntax based on the pretest results. The keywords used to construct research syntaxes were extracted from natural (everyday vocabulary) and controlled (indexer terms) language terms and based on the central concepts associated with the research questions. The researched concepts were suicidal behavior, ICT (AI, machine learning, application, social network, and mobile technology), intervention (information, education, prevention, decision making, and intervention), and program evaluation. The final syntax based on these keywords and used with each database searched is presented in Multimedia Appendix 1. PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore were searched for publications from January 1, 2013, to December 31, 2018. The selection of each database was carefully thought out by the research team (2 senior researchers in suicidology, 2 ICT experts, 1 postdoctoral student, 1 doctoral student, and 1 specialized librarian who provided support) to include data in the medical field, psychology, social and behavioral sciences, and ICT. The number of databases chosen was in line with that of other reviews that focused on specific ICTs and mental health [5,19]. In addition, the choice of databases was based on the following 2 criteria: the number of documents they contain and the degree of overlap of the content indexed in each database [20]. The final choice was made to select high quantities of indexed documents in databases and databases with less content overlap to maximize the completeness of the retrieval process.

Selecting Studies

The final search syntax identified 3848 publications that were imported for EndNote selection. From these, 1391 duplicates were removed. The remaining 2457 publications were then sorted according to the inclusion and exclusion criteria applied by 5 research assistants against titles and abstracts only. Interrater reliability was assessed for 100 publications. Agreement reached 89% and a Cohen κ statistic of 0.704 was achieved, which is considered substantial by Cohen [21]. The studies covered in our review used various methods (qualitative, quantitative, and mixed methods) and were drafted in English or French. They had to provide original empirical or descriptive data. Literature reviews, editorials, theoretical articles, and studies focused only on ethical or legal issues surrounding ICT-based suicide prevention were excluded. The included articles had to directly address suicide prevention strategies using ICT. All studies that used death by suicide, suicidal ideation, or suicide behavior as outcomes were considered. The PRISMA flow chart [16] of our study presented in Figure 1 provides an overview of our study selection process.

XSL•FO
RenderX

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



Charting, Collating, Summarizing, and Reporting Data

A form was developed in Microsoft Excel to retrieve pertinent information from selected qualitative, quantitative, and mixed methods studies to optimize data charting. A research team tested the form before use. It included 10 items: title, objectives, research design, instruments of measure, type of suicide prevention (universal, selective, or indicated), technology category, participant characteristics, results associated with effects and benefits regarding lowering suicidal behavior, clinical and scientific contributions, and recommendations. The research team thoroughly examined 5 independent studies with 10 items and then compared the results. Following this analysis, precision was applied to the exclusion criteria of the studies. Verbal telephone interventions were excluded from this review. Considering the large quantity of data available on suicide prevention telephone interventions [22-24], studies addressing these types of interventions were excluded from this scoping review. However, interventions based on text messaging and smartphone apps were included as new ICT-based interventions relevant to this review. After criteria validation, a summary of pertinent information was prepared for each study included. No systematic methodological quality evaluation was carried out in accordance with the scoping review methodology [11].

Results

General Information

Our search of PubMed, PsycINFO, Sociological Abstracts, and IEEE Xplore yielded 3848 articles. After removing 1391 duplicates and 2265 articles based on a perusal of title and abstract, 192 texts were read in full for the final screening. This allowed us to remove an additional 77 articles, leaving 115 articles for the scoping review. We also noticed through our

XSL•FO
RenderX

analysis that the number of publications increased in 2016 and 2017, and the scientific articles focused more on selective and indicated suicide prevention strategies.

Classification of Interventions' Efficacy

In the context of this scoping review, we aim to provide a general portrait of existing suicide prevention strategies based on ICT. It does not aim to determine the quantitative effectiveness of such interventions. Therefore, a prevention strategy is considered effective when it is based on rigorous theory and when it has been evaluated by a minimum of 2 studies with a quasi-experimental approach [25]. Typically, in traditional systematic reviews, identifying *evidence-based practices* requires a weighting of the methodological quality of primary studies, unlike the scoping review [12]. If the relevant studies are randomized trials, the body of evidence begins with high certainty. If the relevant studies are observational, the body of evidence begins with low certainty [26]. Certain strategies may, however, be presented as *promising* when at least one observational or randomized study supports its efficacy. Therefore, we considered all interventions that have demonstrated effectiveness in any type of research design (qualitative, quantitative, or mixed studies).

Universal Suicide Prevention Strategies

In total, 10 of the studies selected referred to universal suicide prevention strategies (Table 1). These addressed 2 main program categories: (1) health promotion and suicide prevention through the use of educational websites and (2) health promotion and suicide prevention through awareness campaigns and social media psychoeducation. These programs were accessible to all participants. The types of websites identified were information-based, interactive, forums, and chats. Social media platforms included Facebook, Twitter, and personal and professional blogs. In general, the effects of these programs on perceptions and knowledge have been poorly evaluated. Universal suicide prevention strategies included the *SUPREME* (Suicide Prevention through Internet and Media-Based Mental Health Promotion) project [27], the *Storytelling* project [28], the *It Gets Better* project [29], the *Live Through This* project [30], and *Media-Based Prevention Messages* [31]. Although these different universal strategies could have some effects on the negative emotions of participants [32], they primarily play a potential role in reducing the stigma associated with suicide [30,33] and suicidal ideation and behavior [27,30,32] and increasing web-based suicide prevention knowledge [31,34,35], mental health literacy [32,33], and web-based help seeking [33,34]. Table 1 presents a description of the studies that referred to universal suicide prevention strategies.

Table 1. Universal suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Health promotion and suicide prevention through the use of educational websites	[27,28,30,32]	• Improve mental health and well-being • Educate and increase awareness of young people of mental health and suicide prevention • Prevent stigma • Encourage help seeking and service use by young people • Promote protective factors	2	• Websites • Interactive modules • Forums	• Young people • Young adults • At-risk groups	"SUPREME[b]" project [27] and "Live Through This" [30]	Results in terms of reducing suicidal ideation and behavior, increasing mental health literacy, and increasing web-based help seeking seemed promising.
Health promotion and suicide prevention through awareness campaigns and social media psychoeducation	[29,31,33-36]	• Increase social media users' awareness of the existence of and issues associated with suicide and its prevention	4	• Presenting, broadcasting and sharing messages on social media	• Young people • At-risk groups	"It Gets Better" project [29] and "Media-Based Prevention Messages" [31]	Results in terms of reducing suicidal ideation and behavior and improving knowledge, attitudes, and asking for help on the web seemed promising. However, further research was needed.

[a]ICT: information and communication technology.

[b]SUPREME: Suicide Prevention through Internet and Media-Based Mental Health Promotion.

XSL•FO
RenderX

Selective Suicide Prevention Strategies

We identified 53 studies that dealt with selective suicide prevention strategies (Table 2). From these, 9 different selective suicide prevention program categories emerged, and each had multiple different outcomes. These strategies ranged from identifying people at risk to web-based self-management and training programs. Some were designed to be used alone, whereas others were part of a larger intervention program involving direct contact between suicidal individuals and suicide prevention resources. The programs targeted distressed individuals and groups at risk, that is, health professionals. ICT use varied according to the desired outcome. For example, algorithms were used to identify individuals at risk for suicide on the web, as advertised by Google AdWords, which is now known as Google Ads [37,38]; interactive websites were used to support the assessment and management of mental health disorders in collaboration with a mental health professional [39]; and training modules, web-based exercises, and multimedia presentations were used to enhance knowledge, understanding, and attitudes regarding suicide prevention and intervention [40-42]. Programs using algorithms to identify at-risk individuals seemed most promising for identifying people at risk for suicide versus those not at risk [37,38,43-58]. These algorithms analyzed suicide risk on the basis of different elements, including speech and linguistic characteristics, medical notes, search engine ad clicks, and profiles from web-based chat sessions or social media [37,38,43-58]. Although these algorithms helped identify individuals at risk for suicide, a clinical application of these algorithms is yet to be developed, as little is known about the effectiveness of these programs in increasing actual web-based and in-person help-seeking behaviors.

XSL•FO
RenderX

Table 2. Selective suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Identifying at-risk individuals through automatic analysis of their linguistic characteristics during face-to-face consultations	[44-47]	• Identify individuals at risk for suicide and qualify risk level based on speech characteristics	2	• Analysis of characteristics from observation of spoken language	• Individuals presenting risk factors at assessment interviews	Analysis of speech and sound characteristics program [45-47]	This technology seems effective in distinguishing suicidal from nonsuicidal individuals. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of medical file notes and/or research data	[54-57]	• Identify individuals at risk for suicide and qualify risk level based on data (medical file or research data) from clinical observations by clinicians or researchers	1	• Analysis of written language	• Individuals presenting risk factors during assessment interviews	The "SafetyNet" program [56]	This technology seems effective in distinguishing suicidal from nonsuicidal individuals. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of their linguistic characteristics during chat sessions or on written forms	[43,58]	• Identify individuals at high risk for suicide based on characteristics of their words or responses	0	• Analysis of written language	• Individuals presenting risk factors during written exchanges	Suicide meaning making applied in specific contexts [58]	This technology seems promising. A clinical use of this technology is yet to be developed.
Identifying at-risk individuals through automatic analysis of their linguistic characteristics on social media	[48-53,59-62]	• Examine writings of social media users to automatically identify individuals at risk for suicide and offer them proactive help	4	• Analysis of written language	• Social media users	Program to screen suicidal individuals based on tweets [52,59]	This technology is innovative. Further research is needed.
Identifying at-risk individuals through targeted ads on search engines	[37,38]	• Identify suicidal individuals through their queries on search engines • Suggest help resources	2	• Search engine algorithm and targeted ads	• General public	Web-based sentinel program [37]	The number of clicks on targeted ads serves as an indicator. There are too few results to date to evaluate these programs' effectiveness in increasing the web-based help-seeking behaviors of suicidal individuals.
Web-based identification programs tailored to different groups	[63-69]	• Identify vulnerable groups using mailing lists or websites to direct them toward help resources	3	• Web-based standardized risk assessment questionnaires with response algorithms adapted to assessment results	• Individuals at risk for suicide	EMPATHY[b] [63] and HEAR[c] for nurses [69]	This type of program allows identifying individuals at risk who have no contact with health services. Utilization rates for proposed resources seem encouraging.

XSL•FO
RenderX

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Web-based mental health self-evaluation programs	[39,70]	• Support evaluation and management of mental health problems in conjunction with a mental health professional	1	• Interactive website • Web-based evaluation tool • Algorithm tool for redirecting to resources • Development of a web-based personalized intervention plan	• Individuals receiving support from a health professional	myGRaCE decision support system [39] and YouthCHAT [70]	Improves patient engagement in understanding their situation and in managing their treatment.
Web-based mental health self-management program with measures of impact on suicide risk	[71-81]	• Facilitate web-based mood self-management and improve quality of life through a cognitive behavioral intervention	5	• Interactive website • Psychoeducation module • Multimedia presentation • Web-based exercises • Discussion and consultation forum	• Individuals with low-intensity mental health problems	MindSpot Clinic [77]	Program membership rates are relatively low. Individuals tend more often to experience a decrease in suicidal ideation and depression symptoms. Programs are offered alone or in conjunction with clinical follow-up by a professional. Results are hard to compare.
Web-based suicide prevention training programs	[40-42,82-87]	• Improve knowledge, attitudes and suicide prevention practices of professionals and sentinels	5	• Website • Training modules • Multimedia demonstration • Web-based exercises	• Professionals • Sentinels	Question, Persuade, Refer, and Treat program [40,84]	These programs allow improving knowledge and attitudes. They seem less effective in changing intervention practices if not offered along with certain forms of face-to-face practices.

[a]ICT: information and communication technology.

[b]EMPATHY: Empowering a Multimodal Pathway Toward Healthy Youth.

[c]HEAR: Healer Education Assessment and Referral.

Using ICT to identify individuals at risk for suicide in school (from elementary school to medical school) or community settings has been associated with increased mental health literacy, help-seeking behavior, and use of help resources [63-65]. For example, in the Health Education Assessment and Referral (HEAR) program, medical students anonymously responded to a web-based questionnaire on suicide risk and various mental health issues. The program allowed identifying students at risk for suicide and referring them to a medical school psychiatrist or psychologist through a web-based platform. As a result, the use of medical schools' mental health services increased from 11.5% to 15% over 4 years [64].

Our review also identified web-based self-assessment and self-management programs. For example, myGRaCE is a decision-making support system that combines service user self-assessment and practitioner expertise by comparing the user's self-assessment against a practitioner's assessment [39]. Most of the participants in this study agreed that myGRaCE helped them assess their personal security, understand what puts them in danger, and what areas they should change in their lives. Self-assessment programs are also a way to engage young people and adults by helping them understand their situation and identify ways to take control of their health [39,70]. As for web-based self-management programs, results showed that many programs could, in some cases, significantly reduce suicidal ideation [71-74] and increase chances of resorting to mental health interventions [75,76]. However, the rate of adherence to these types of programs remains low. Examples of these include MoodGYM [71], MindSpot Clinic [77], different types of internet-based cognitive behavioral therapy (iCBT), Thrive [73], the Sadness program [78], and CATCH-IT (Competent Adulthood Transition with Cognitive, Behavioral,

XSL•FO
RenderX

Humanistic and Interpersonal Training) [74]. These programs were offered either alone or in combination with clinical follow-up. Users seemed to appreciate the mobile apps used for self-management and suicide prevention training. Many web-based suicide prevention programs have been tested for their efficacy. These programs were intended for gatekeepers who work with adolescents [82], gatekeepers in school settings [83], mental health professionals [40], health professionals in general [41,42,84], Veterans Affairs providers [85], and graduate students [86]. In some cases, face-to-face training developed more knowledge and suicide prevention skills than web-based training [84]. In addition, an increase in suicide prevention knowledge was often observed within the first months post training but often decreased over time [82]. Details of the studies that referred to selective suicide prevention strategies are presented in Table 2.

Indicated Suicide Prevention Strategies

Regarding the indicated suicide prevention strategies, we selected 52 studies for our review and identified 9 program categories. These programs, presented in Table 3, were offered by health professionals or psychosocial help services and covered suicide risk assessment triage and monitoring, crisis intervention, low-intensity psychological interventions, psychotherapy for individuals at risk for suicide, and technological tools to support face-to-face interventions. The ICTs used for these programs included classification algorithms, chats and text messages, mobile apps, and interactive clinical intervention websites. Examples of these programs include the Mental Health eClinic [55], a web-based assessment and self-management program; Reframe-IT [88], an internet-based CBT program for high school students at risk for suicide; MYPLAN [89], a mobile phone safety plan app for supporting people at risk for suicide; dialectical behavioral therapy Coach Mobile [90], a CBT support intervention mobile phone app; and RAFT (Reconnecting After a Suicide Attempt) [91], a brief web-based outreach intervention for people who attempted suicide and who lost contact with services. In general, ICT-based self-assessment tools have been reported to be as effective in person as in hardcopy format in identifying suicide risk warning signs [92].

XSL•FO
RenderX

Table 3. Indicated suicide prevention strategies.

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Development of web-based suicide risk assessment tools	[92-97]	• Propose suicide risk assessment tools administered electronically in clinical contexts	4	• Electronic clinical instrument • Response selection algorithm based on assessment results	• Individuals at risk for suicide in contact with health services	Web-based Columbia Suicide Severity Rating Scale [93]	Electronic assessment tools seem as effective as hardcopy versions. They allow internet users at times to reveal their suicidal behaviors more easily
Use of artificial intelligence and machine learning to optimize completion time for suicide risk assessment tools	[98]	• Reduce completion time for suicide risk assessment tools	1	• CAT[b]	• Individuals at risk for suicide in contact with health services	CAT [98]	This type of algorithm allows reducing the number of items needed to assess suicide risk.
Use of triage systems by ICT to assess suicide risk	[99]	• Improve triage of individuals at risk for suicide who use health services	1	• On-lining of clinical suicide risk assessment tools • Risk-level classification algorithm based on scores • Targeted offer of resources • Automatic alert to the clinical team	• Individuals at risk for suicide in contact with health services	Mental health eClinic [99]	This type of program allows identifying young people at risk for suicide and offering them clinical treatment more rapidly.
Crisis intervention via text messaging or internet chatting	[68,69,84,85]	• Intervene in a crisis situation via text messaging or internet chatting	3	• Text messaging • Internet chatting • Computerized system for processing text messages and chat exchanges	• Young people are the priority target group of these interventions. They can, however, be used to reach other age groups	Kids Help Phone LIVECHAT Project [100] and RAFT[c] [91]	Text messages and internet chat interventions seem a viable alternative to telephone interventions for certain groups. However, intervention strategies adapted to these modes of communication need to be developed.
Use of web-based publications by patients as intervention support material	[101,102]	• Assess suicide risk and intervene using a person's web-based discourse	1	• Content of posts written on social networks	• Suicidal individuals receiving mental health services	Patient's social networking sites as a clinical tool [101]	The use of content written on social networks by patients allows a better understanding of the situation and helps perform a suicide risk assessment, especially in the case of patients who deny such behaviors.

XSL•FO
RenderX

Type of program	Identified studies	Objectives	Programs, n	ICT[a] used	Targeted population	Examples	Results and comments
Web-based management and low-intensity intervention for individuals at risk for suicide	[1,31,63,68,88,102-114]	• Intervene with individuals at risk for suicide using a cognitive behavioral approach	7	• Interactive website • Web-based exercise module • Discussion forum • Multimedia presentation	• General population • Individuals referred to programs by clinicians	Reframe-IT [88], Fitmindkit [102], EMPATHY[d] [63], and SMART[e] Mental Health Project [110,113]	These programs bring about a small decrease in suicide risk among participants.
Mobile apps used as part of treatment follow-up with suicidal individuals	[90,115-121]	• Support face-to-face clinical intervention • Support aboriginal young people by complementing face-to-face intervention	2	• Interactive mobile app • Algorithm for providing responses and offering resources based on information provided by the user	• Individuals receiving mental health services • Aboriginal young people	DBT[f] Coach Mobile [90] and AIMhi Stay Strong iPad [101]	These apps seem to help reduce the danger of suicide and the urgency of self-harm. Perceptions are positive, and apps allow reaching young people in isolated communities.
Use of web-based monitoring tools to improve psychological follow-up of individuals at risk for suicide	[107,122,123]	• Conduct regular evaluations of individuals and their symptoms to inform and adjust intervention	3	• Digitalization of screening tools • Mobile app • Contact with intervention team via automated alerts • Automated sending of text messages	• Individuals diagnosed with a mood disorder and receiving mental health services	Depression Project [123]	The use of an app facilitates disclosure of suicidal and self-harm behaviors. Regular monitoring of symptoms helps adjust intervention strategies.
Follow-up program by automatic text messages for individuals at risk for suicide	[91,124-126]	• Offer tailored follow-up following a suicide attempt to increase treatment use and reduce suicidal behaviors and self-harm	3	• Automated sending of predrafted text messages, including encouragement and appointment date reminders	• Suicidal individuals receiving mental health services	Postattempt follow-up program by text messaging [124]	Users appreciate the text messages and deem them a good way of keeping in touch with care services. Help-seeking behaviors increase, and self-harm behaviors decrease.

[a]ICT: information and communication technology.

[b]CAT: computerized adaptive testing.

[c]RAFT: Reconnecting After a Suicide Attempt.

[d]EMPATHY: Empowering a Multimodal Pathway Toward Healthy Youth.

[e]SMART: Systematic Medical Appraisal, Referral and Treatment.

[f]DBT: dialectical behavioral therapy.

In some cases, web-based self-assessment tools, as opposed to face-to-face tools, seemed to facilitate the disclosure of suicidal behavior [93]. AI has also been shown to be effective in optimizing the web-based assessment of suicide behavior by using an item response-based computer-adaptive simulation to reduce the length of a suicide risk assessment tool [98]. As for ICT-based triage systems, the use of the Synergy Online System allowed young people to complete a web-based clinical

XSL•FO
RenderX

assessment before a face-to-face or web-based clinical appointment, who were to be prioritized and contacted immediately in case of high suicide risk [99]. Web-based crisis interventions by chat or SMS text messaging have been on the rise and seem to be an alternative to phone interventions for some groups. Studies comparing chat services with SMS text messaging have shown similar results [127]. However, users of chat services, such as 113Online [128], were more likely to be at a high level of suicide crisis, had more mental health problems, were younger, and were more often to be women compared with crisis hotline users. In addition, interventions delivered through chat services were longer and more complicated, and fewer changes were observed in the individual's emotional state [127,128]. Interventions have yet to be developed and adapted to this type of technology; the fact that interventions designed to be delivered by telephone were delivered by chat or SMS text messaging without being adapted to these other media was considered a major limitation [128]. Some studies also analyzed content posts by suicidal patients on social networking sites to better assess suicide risk, especially in patients denying suicidal behavior [101,129]. Although social media content helped with better understanding a person's situation, there were many ethical concerns regarding this data collection method [129].

Self-management and iCBT programs seemed to reduce suicidal ideation. Unlike iCBT and self-management interventions addressing general mental health in the selective suicide prevention section, these interventions specifically addressed user suicide behaviors. Examples of these programs included Empowering a Multimodal Pathway Toward Healthy Youth [63], Reframe-IT [88], Fitmindkit [102], Safe conversation [103], Latitudes [103] and PrevenDep [104]. Intervention goals included developing basic problem-solving skills [88] and cognitive restructuring [105] to reduce suicide risk. Other ICT suicide prevention interventions include using computer alerts or mobile apps to apply a safety plan. These ranged from a system alerting the clinician not to forget to use the safety plan with a patient [130] to a mobile app of a personalized safety plan that patients could have on their phones at all times [89,115]. Research has demonstrated a good level of acceptability [131] for these interventions, but further evaluation of their effects on suicidal behaviors is necessary. Mobile apps have also been used to support therapeutic follow-up for suicidal patients [90,115,116], including for specific groups such as Aboriginals and Torres Strait Island Australians [117]. These mobile apps seemed to help reduce suicidal danger and self-harm [90,118], reach young people in isolated communities [132], and increase adherence to face-to-face follow-ups [119].

Discussion

Limitations

The rapid expansion of ICT use in suicide prevention has preceded the development of theoretical models to orient its methods and content and the accumulation of sufficient empirical research to indicate what is most helpful and what is not helpful. In this context, common sense has been the guiding principle, resulting in a plethora of suicide prevention activities being implemented with limited research evaluations of their effectiveness slowly following. Many of the most widely used ICT suicide prevention services have never been evaluated. Therefore, the published studies concern a small nonrandom sample of select interventions that a few researchers have been interested in studying. When little or no research has been published on ICT programs, this does not mean that they are not effective. Similarly, when there are promising empirical data on the benefits of a program, this does not mean that it is better or more useful than programs that have not been studied. Furthermore, we do not know if the *promising* results reported will stand the test of time, as more research is conducted using more rigorous research methodologies. Therefore, any conclusions drawn from the limited scientific publications must be considered preliminary and hopefully will be subject to verification in the future.

Contributions

This scoping review on ICT use in suicide prevention shows that a large number of studies have been published in the past few years [1,3,5]. Our findings shed light on the use of ICT in different types of universal, selective, and indicated suicide prevention strategies.

Amid publications with the potential to reach a large population, there are only a few publications on the use of ICT in universal suicide prevention [27-35,133]. Our findings reveal that there are around 85% fewer studies on universal prevention strategies than those on selective and indicated strategies. The 10 publications we identified describe health promotion and suicide prevention through educational websites, web-based awareness campaigns, and social media [27-35,133]. The limited empirical findings suggest that these programs may play a role in increasing general mental health literacy and the incidence of help-seeking behavior, which are associated with reduced suicidal ideation and behavior. However, the effects of these programs on perceptions and knowledge have rarely been investigated. It is important to identify which characteristics of educational websites, awareness campaigns, and postings on social media are associated with positive changes in help seeking and reductions in suicidal ideations and behaviors. Furthermore, it is essential to develop effective means for identifying sites and postings that are helpful and notifying users about or orienting them toward internet content that may be of help to them.

Selective suicide prevention strategies using ICT consist of programs that identify specific subpopulations to offer specialized support to reduce suicide risk. They particularly targeted young people and various at-risk groups (eg, the lesbian, gay, bisexual, transgender, queer, and other community; sexual minorities; and aboriginal communities) through their profiles and linguistic characteristics (eg, chat sessions, web-based written forms, and social media) or by analyzing their medical records and research data. They also include targeted web-based advertising, web-based self-assessment and self-management, and web-based training. A larger published body of research indicates that these approaches are promising, particularly in controlled environments such as schools. School programs using ICT, such as HEAR, have been shown to have a positive impact

XSL•FO
RenderX

on mental health help seeking and service use in youth [64]. Programs targeting specific subgroups of the general population have been shown to increase mental health literacy and the chances of using mental health services, but this effect seems to decrease over time [71,73,74,77,92]. Therefore, it is important to focus on the sustainability of the effects of programs that have an initial positive impact. The novelty of an intervention may be associated with greater effects. If novelty is a key feature of programs' success, then either the programs need to be constantly changed and renewed to sustain their impact or the programs need to be continually replaced by new and different activities to ensure that people will continue to be helped over time. Therefore, both approaches may be proven to be unsustainable over time. This rapid effect can also give a false sense that resorting to web-based strategies is sufficient. However, web-based help does not replace face-to-face or direct support and care from trained professionals. These different selective suicide prevention strategies should only be used in addition to face-to-face type of help and interventions. Moreover, our empirical findings show that more research is required to better distinguish between false positives and false negatives in these web-based identification techniques. It is of utmost importance to help identify at-risk groups and avoid discarding individuals who are assessed as false negatives too quickly. A false impression of security can put these individuals at a greater risk.

Where indicated suicide prevention strategies are concerned, programs vary, including using ICT in suicide risk assessment triage, monitoring suicide risk, crisis intervention activities, and psychotherapy. The programs were offered on the web only (eg, websites, mobile phone apps, and chats), but they were sometimes supported by face-to-face interventions. In some instances, they were found to be efficient (eg, shortened web-based assessments), and sometimes they were time-consuming (eg, interventions were longer in chat sessions). ICT-based and self-management interventions seem promising, as they address suicidal behaviors directly and offer alternatives for coping with suicidal thoughts by developing problem-solving skills [88,90-92]. In addition, specific at-risk groups in isolated communities where other services are not available may benefit from mobile apps used for patient follow-up.

Ethical Considerations to Address in the Future

Many of the studies included in this review raised security and ethical concerns regarding web-based suicide prevention practices. Ethical concerns range from a lack of training and web-based moderators' skills to the lack of an evidence-based framework providing guidelines for the secure use of ICT [35]. As mentioned by Robinson et al [35], it is ethically necessary to provide security protocols and a clear code of ethics for safe web-based intervention. Another ethical concern is that all individuals have the right to privacy, including web privacy [101,129]. Thus, there is reason to question whether the content analysis of web-based social media posts, emails, and other web-based sources of information is an ethical research endeavor if informed consent is not obtained from the individuals who have posted the information [101,129]. Beyond security and privacy, there are many other ethical concerns regarding web-based surveillance, informed consent, communication, controls, and disclosure [134,135]. For example, a proper ethical assessment of risks and benefits to the use of different ICT strategies is rarely, if almost never, considered in the development process. There is also a major ethical concern regarding reducing or eliminating in-person services and replacing them with insufficient web-based solutions that may appear to present better cost-effectiveness without proper in-depth assessment. With this concern, web-based interventions should always be combined with in-person formal help and intervention services. This raises many concerns and provides plenty of grounds for further research aimed at ensuring safer and more ethical use of ICT in suicide prevention.

Conclusions

As the number of studies on ICT use in suicide prevention is growing, the published literature needs to be reviewed regularly. This scoping review shows that ICT use in suicide prevention provides an interactive, personalized, readily available, and accessible approach to reach various populations for identification of at-risk individuals and to provide support. ICT may provide a sense of *being connected* to people who are otherwise isolated and reluctant to use offline services. Promising published findings on web-based intervention content includes psychoeducation and skills training. As digital help proliferates, one should consider whether this means that help-seeking and suicide prevention activities will replace traditional offline services. However, in some areas where radical changes were expected, existing modalities continue to be used (eg, individuals did not stop using in-person services when telephone phone crisis lines appeared, and these telephone lines reached a different audience and complemented existing services). The extent to which ICT will become the main source of suicide prevention activities will depend upon its efficacy in helping people, compared with and as a complement to existing services and activities.

Although there is a growing body of evidence regarding ICT use in suicide prevention, program evaluation is still lacking. There is a need for more research evaluating and comparing the impacts of various ICT strategies in different contexts, understanding the profiles of individuals at risk of suicide who use ICT, and web-based help-seeking behaviors. We also need to better understand the impact of ICT on individuals who are bereaved by suicide. Moreover, ethics and security concerns regarding web-based suicide prevention have been the focus of very limited research and need to be addressed in future studies.

Furthermore, service users, providers, and managers from private or public systems should be informed as of now and updated regularly on the benefits and risks of ICT use in health and social services. This information should include (1) effectiveness in various well-described contexts and potential unexpected iatrogenic effects, (2) cost-benefit relationship to the best comparator, (3) access, (4) acceptability and ethical concerns, (5) security, and (6) implementation. Concerning implementation, quality standards should be similar to those of Improving Access to Psychological Therapies standards set by the United Kingdom for access to psychotherapy. These standards are (1) a model of care, (2) access, (3) evidence-based interventions, (4) outcome-based measurement, and (5) the

XSL•FO
RenderX

provider's training and supervision. International health and social technology assessment agencies should therefore consider developing guidelines and a system of voluntary accreditation for ICT use in suicide prevention. National and regional public health and social services may require that, before commissioning an ICT for suicide prevention or mental health care, their Health Technology Agency recommends ICT based on its efficacy, efficiency, safety, acceptability, and feasibility in the context of jurisdiction.

Acknowledgments

The authors would like to thank Brian Mishara, Alain Lesage, Jérôme Gaudreault, and the Association Québécoise de prevention du suicide for their valuable and constructive suggestions during the planning and development of this study. A grant was awarded for the completion of the study to RL and CB by the Canadian Institutes of Health Research, Institute of Population and Public Health (grant 385 708). JR received a postdoctoral grant from the Quebec Network on Nursing Intervention Research and the project from the Quebec Network on Suicide, Mood Disorders and Related Disorders.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy syntax specific to each consulted database.
[DOCX File , 15 KB - jmir_v23i5e25288_app1.docx]

References

1. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
2. van Spijker BA, Werner-Seidler A, Batterham PJ, Mackinnon A, Calear AL, Gosling JA, et al. Effectiveness of a web-based self-help program for suicidal thinking in an Australian community sample: randomized controlled trial. J Med Internet Res 2018 Feb 14;20(2):e15 [FREE Full text] [doi: 10.2196/jmir.8595] [Medline: 29444769]
3. Perry Y, Werner-Seidler A, Calear A, Christensen H. Web-based and mobile suicide prevention interventions for young people: a systematic review. J Can Acad 2016;25(2):73-79. [Medline: 27274742]
4. de la Torre I, Castillo G, Arambarri J, López-Coronado M, Franco MA. Mobile apps for suicide prevention: review of virtual stores and literature. JMIR mHealth and uHealth 2017 Oct 10;5(10):e130. [doi: 10.2196/mhealth.8036] [Medline: 29017992]
5. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
6. Robinson J, Bailey E, Witt K, Stefanac N, Milner A, Currier D, et al. What works in youth suicide prevention? A systematic review and meta-analysis. EClinicalMedicine 2018;4-5:52-91 [FREE Full text] [doi: 10.1016/j.eclinm.2018.10.004] [Medline: 31193651]
7. van Spijker BAJ, van Straten A, Kerkhof AJ. Effectiveness of online self-help for suicidal thoughts: results of a randomised controlled trial. PLoS ONE 2014 Feb 27;9(2):e90118. [doi: 10.1371/journal.pone.0090118]
8. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
9. Organisation mondiale de la Santé. Prévention du suicide: l'état d'urgence mondial. Geneva: Organisation mondiale de la Santé; 2014.
10. Le grand dictionnaire terminologique (GDT). Office québécois de la langue française. URL: http://gdt.oqlf.gouv.qc.ca/ [accessed 2020-09-16]
11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
12. Levac D, Colquhoun H, O'Brien K. Scoping studies: advancing the methodology. Implementation Science 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69]
13. Mishara B, Kerkhof AJFM, editors. Suicide Prevention and New Technologies: Evidence-Based Practice. London: Palgrave MacMillan; 2013.
14. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
15. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]

XSL•FO
RenderX

16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, the PRISMA Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;349(jan02 1):g7647-g7647 [FREE Full text] [doi: 10.1136/bmj.g7647]
17. Methely A, 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 Services Research 2014;14(1):1-10 [FREE Full text] [doi: 10.1186/s12913-014-0579-0]
18. Straus SE, Tetroe J, Graham ID, Zwarenstein M, Bhattacharyya O, Shepperd S. Monitoring use of knowledge and evaluating outcomes. Canadian Medical Association Journal 2010 Jan 18;182(2):E94-E98. [doi: 10.1503/cmaj.081335]
19. Pourmand A, Roberon J, Caggiula A, Monsalve N, Rahimi M, Torres-Llenza V. Social media and suicide: a review of technology-based epidemiology and risk assessment. Telemedicine and e-Health 2019;25(10):880-888 [FREE Full text] [doi: 10.1089/tmj.2018.0203]
20. Bartels EM. How to perform a systmatic search. Best Practice & Research Clinical Rheumatology 2013;27(2):295-306 [FREE Full text] [doi: 10.1016/j.berh.2013.02.001]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
22. Mishara B, Chagnon F, Daigle M, Balan B, Raymond S, Marcoux I, et al. Comparing models of helper behavior to actual practice in telephone crisis intervention: A silent monitoring study of calls to the US 1-800-SUICIDE network. Suicide and Life-Threatening Behavior 2007;37(3):291-307. [doi: 10.1521/suli.2007.37.3.291]
23. Mishara B, Daigle M. Helplines crisis intervention services: challenges for the future. In: Lester D, editor. Suicide Prevention: resources for the millennium. New York: Routledge; 2001:153-171.
24. Mishara BL, Daigle M, Bardon C, Chagnon F, Balan B, Raymond S, et al. Comparison of the Effects of Telephone Suicide Prevention Help by Volunteers and Professional Paid Staff: Results from Studies in the USA and Quebec, Canada. Suicide Life Threat Behav 2016 Oct;46(5):577-587 [FREE Full text] [doi: 10.1111/sltb.12238] [Medline: 26946110]
25. Henschke N, Ostelo R, Vlaeven S, Morley S, Assendelft W, Main C. Behavioural treatment for chronic low-back pain. Cochrane Database System Reviews.: Cochrane Library; 2010. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7065591/pdf/CD002014.pdf [accessed 2020-09-15]
26. Schünemann H, Vist G, Higgins J, Santesso N, Deeks J, Glasziou P, et al. Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Chichester (UK): John Wiley & Sons; 2019:403-431.
27. Carli V. Preventing suicidality through online tools: The SUPREME Project. In: Courtet P, editor. Understanding Suicide: from Diagnosis to Personalized Treatment. New York: Springer; 2016:281-289.
28. Wexler L, Gubrium A, Griffin M, DiFulvio G. Promoting Positive Youth Development and Highlighting Reasons for Living in Northwest Alaska Through Digital Storytelling. Health Promotion Practice 2012 Oct 24;14(4):617-623. [doi: 10.1177/1524839912462390]
29. Hurley, SJ. Public pedagogy and the experience of video creators in the It Gets Better Project. Libraries digital conservancy. 2014. URL: https://hdl.handle.net/11299/171186 [accessed 2020-09-16]
30. Rogers ML, Schneider ME, Gai AR, Gorday JY, Joiner TE. Evaluation of two web-based interventions in reducing the stigma of suicide. Behaviour Research and Therapy 2018 Oct;109:49-55. [doi: 10.1016/j.brat.2018.08.001]
31. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing Social Media-Based Suicide Prevention Messages in Partnership With Young People: Exploratory Study. JMIR Ment Health 2017 Oct 04;4(4):e40 [FREE Full text] [doi: 10.2196/mental.7847] [Medline: 28978499]
32. Till B, Tran US, Voracek M, Niederkrotenthaler T. Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno . Werther effects. Br J Psychiatry 2018 Jan 02;211(2):109-115. [doi: 10.1192/bjp.bp.115.177394]
33. Taylor-Rodgers E, Batterham PJ. Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. J Affect Disord 2014 Oct 15;168:65-71. [doi: 10.1016/j.jad.2014.06.047] [Medline: 25038293]
34. Berger J. Adolescents on the lookout for suicidal friends on social networking sites. Digital Repository at the University of Maryland.: Doctoral dissertation, University of Maryland, College Park; 2013. URL: https://drum.lib.umd.edu/handle/1903/14494 [accessed 2020-09-22]
35. Robinson J, Rodrigues M, Fisher S, Bailey E, Herrman H. Social media and suicide prevention: findings from a stakeholder survey. Shanghai Arch Psychiatry 2015 Feb 25;27(1):27-35 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.214133] [Medline: 25852253]
36. Chen JI, Smolenski DJ, Dobscha SK, Bush NE, Denneson LM. Correlates of Mental Health Smartphone Application Use among Patients with Suicidal Ideation. Journal of Technology in Human Services 2018 Nov 16;36(4):191-207. [doi: 10.1080/15228835.2018.1502119]
37. Sueki H, Ito J. Suicide prevention through online gatekeeping using search advertising techniques: a feasibility study. Crisis 2015 Jul;36(4):267-273. [doi: 10.1027/0227-5910/a000322] [Medline: 26440623]

XSL•FO
RenderX

38. Liu NH, Contreras O, Muñoz RF, Leykin Y. Assessing Suicide Attempts and Depression Among Chinese Speakers Over the Internet. Crisis 2014 Sep 01;35(5):322-329. [doi: 10.1027/0227-5910/a000261]
39. Buckingham CD, Adams A, Vail L, Kumar A, Ahmed A, Whelan A, et al. Integrating service user and practitioner expertise within a web-based system for collaborative mental-health risk and safety management. Patient Education and Counseling 2015 Oct;98(10):1189-1196. [doi: 10.1016/j.pec.2015.08.018]
40. Gryglewicz K, Chen JI, Romero GD, Karver MS, Witmeier M. Online Suicide Risk Assessment and Management Training. Crisis 2017 May;38(3):186-194. [doi: 10.1027/0227-5910/a000421]
41. Ryan K, Tindall C, Strudwick G. Enhancing Key Competencies of Health Professionals in the Assessment and Care of Adults at Risk of Suicide Through Education and Technology. Clinical Nurse Specialist 2017;31(5):268-275. [doi: 10.1097/nur.0000000000000322]
42. LeCloux M. The development of a brief suicide screening and risk assessment training webinar for rural primary care practices. Journal of Rural Mental Health 2018 Jan;42(1):60-66. [doi: 10.1037/rmh0000087]
43. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu M, Zhang XY, et al. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. J Affect Disord 2016 Mar 15;193:109-116 [FREE Full text] [doi: 10.1016/j.jad.2015.12.066] [Medline: 26773901]
44. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. Computational and Mathematical Methods in Medicine 2016;2016:1-8. [doi: 10.1155/2016/8708434]
45. Pestian J, Grupp-Phelan J, Bretonnel CK, Meyers G, Richey L, Matykiewicz P, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide & Life-Threatening Behavior 2016;46:154-159 [FREE Full text] [doi: 10.1111/sltb.12180]
46. Pestian J, Sorter M, Connolly B, Cohen K, McCullumsmith C, Gee J, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide and Life-Threatening Behavior 2017;47:112-121. [doi: 10.1111/sltb.12312]
47. Venek V, Scherer S, Morency L, Rizzo AS, Pestian J. Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction. IEEE Trans. Affective Comput 2017 Apr 1;8(2):204-215. [doi: 10.1109/taffc.2016.2518665]
48. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. JMIR Ment Health 2016 May 16;3(2):e21 [FREE Full text] [doi: 10.2196/mental.4822] [Medline: 27185366]
49. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. J Med Internet Res 2017 Jul 10;19(7):e243. [doi: 10.2196/jmir.7276]
50. Guan L, Hao B, Cheng Q, Yip PS, Zhu T. Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model. JMIR Mental Health 2015 May 12;2(2):e17. [doi: 10.2196/mental.4227]
51. Lv M, Li A, Liu T, Zhu T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. PeerJ 2015;3:e1455 [FREE Full text] [doi: 10.7717/peerj.1455] [Medline: 26713232]
52. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A Linguistic Analysis of Suicide-Related Twitter Posts. Crisis 2017;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
53. Ren F, Kang X, Quan C. Examining Accumulated Emotional Traits in Suicide Blogs With an Emotion Topic Model. IEEE J. Biomed. Health Inform 2016 Sep;20(5):1384-1396. [doi: 10.1109/jbhi.2015.2459683]
54. Hettige NC, Nguyen TB, Yuan C, Rajakulendran T, Baddour J, Bhagwat N, et al. Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach. General Hospital Psychiatry 2017 Jul;47:20-28. [doi: 10.1016/j.genhosppsych.2017.03.001]
55. Batterham PJ, van Spijker BAJ, Mackinnon AJ, Calear AL, Wong Q, Christensen H. Consistency of trajectories of suicidal ideation and depression symptoms: Evidence from a randomized controlled trial. Depress Anxiety 2018 Nov 22;36(4):321-329. [doi: 10.1002/da.22863]
56. Roaten K, Johnson C, Genzel R, Khan F, North CS. Development and Implementation of a Universal Suicide Risk Screening Program in a Safety-Net Hospital System. The Joint Commission Journal on Quality and Patient Safety 2018 Jan;44(1):4-11. [doi: 10.1016/j.jcjq.2017.07.006]
57. Adamou M, Antoniou G, Greasidou E, Lagani V, Charonyktakis P, Tsamardinos I, et al. Toward Automatic Risk Assessment to Support Suicide Prevention. Crisis 2019 Jul;40(4):249-256. [doi: 10.1027/0227-5910/a000561]
58. Silvén Hagström A. 'Suicide stigma' renegotiated: Storytelling, social support and resistance in an Internet-based community for the young suicide-bereaved. Qualitative Social Work 2016 Apr 27;16(6):775-792. [doi: 10.1177/1473325016644039]
59. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001]
60. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, et al. Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Inform Decis Mak(Suppl 2) 2018;18:43 [FREE Full text] [doi: 10.1186/s12911-018-0632-8]
61. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed Inform Insights 2018 Aug 27;10:117822261879286. [doi: 10.1177/1178222618792860]

XSL•FO
RenderX

62. Jung H, Park H, Song T. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. J Med Internet Res 2017 Jul 24;19(7):e259. [doi: 10.2196/jmir.7452]
63. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Long-term Results from the Empowering a Multimodal Pathway Toward Healthy Youth Program, a Multimodal School-Based Approach, Show Marked Reductions in Suicidality, Depression, and Anxiety in 6,227 Students in Grades 6–12 (Aged 11–18). Front. Psychiatry 2017 May 15;8:81. [doi: 10.3389/fpsyt.2017.00081]
64. Downs N, Feng W, Kirby B, McGuire T, Moutier C, Norcross W, et al. Listening to Depression and Suicide Risk in Medical Students: the Healer Education Assessment and Referral (HEAR) Program. Acad Psychiatry 2014 Apr 5;38(5):547-553. [doi: 10.1007/s40596-014-0115-x]
65. Haskins J, Carson JG, Chang CH, Kirshnit C, Link DP, Navarra L, et al. The Suicide Prevention, Depression Awareness, and Clinical Engagement Program for Faculty and Residents at the University of California, Davis Health System. Acad Psychiatry 2015 Jun 11;40(1):23-29. [doi: 10.1007/s40596-015-0359-0]
66. Williams D. Consensus-based strategies for using social networking to reduce suicide risk: a Delphi study. In: Doctoral Dissertation. Denver: The University of the Rockies; 2016.
67. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):630-636 [FREE Full text] [doi: 10.1037/a0038805] [Medline: 25688811]
68. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Initial Findings from a Novel School-Based Program, EMPATHY, Which May Help Reduce Depression and Suicidality in Youth. PLoS ONE 2015 May 14;10(5):e0125527. [doi: 10.1371/journal.pone.0125527]
69. Davidson JE, Zisook S, Kirby B, DeMichele G, Norcross W. Suicide Prevention: a healer education and referral program for nurses. JONA: The Journal of Nursing Administration 2018;48(2):85-92. [doi: 10.1097/nna.0000000000000582]
70. Goodyear-Smith F, Martel R, Darragh M, Warren J, Thabrew H, Clark TC. Screening for risky behaviour and mental health in young people: the YouthCHAT programme. Public Health Rev 2017 Oct;38:20 [FREE Full text] [doi: 10.1186/s40985-017-0068-1] [Medline: 29450092]
71. Christensen H, Farrer L, Batterham PJ, Mackinnon A, Griffiths KM, Donker T. The effect of a web-based depression intervention on suicide ideation: secondary outcome from a randomised controlled trial in a helpline. BMJ Open 2013 Jun 28;3(6):e002886. [doi: 10.1136/bmjopen-2013-002886]
72. Mewton L, Andrews G. Cognitive behaviour therapy via the internet for depression: a useful strategy to reduce suicidal ideation. J Affect Disord 2015 Jan 01;170:78-84. [doi: 10.1016/j.jad.2014.08.038] [Medline: 25233243]
73. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014 May;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
74. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized Clinical Trial of a Primary Care Internet-based Intervention to Prevent Adolescent Depression: One-year Outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
75. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-Based Cognitive Behavioral Therapy Intervention for the Prevention of Suicidal Ideation in Medical Interns: A Randomized Clinical Trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]
76. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. Journal of Affective Disorders 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023]
77. Nielssen O, Dear BF, Staples LG, Dear R, Ryan K, Purtell C, et al. Procedures for risk management and a review of crisis referrals from the MindSpot Clinic, a national service for the remote assessment and treatment of anxiety and depression. BMC Psychiatry 2015;15:304 [FREE Full text] [doi: 10.1186/s12888-015-0676-6] [Medline: 26626712]
78. Williams AD, Andrews G. The Effectiveness of Internet Cognitive Behavioural Therapy (iCBT) for Depression in Primary Care: A Quality Assurance Study. PLoS ONE 2013 Feb 22;8(2):e57447. [doi: 10.1371/journal.pone.0057447]
79. Lara MA, Tiburcio M, Aguilar AA, Sánchez-Solís A. A four-year experience with a Web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica 2014;35(5-6):399-406 [FREE Full text] [Medline: 25211568]
80. Gleeson J, Lederman R, Wadley G, Bendall S, McGorry P, Alvarez-Jimenez M. Safety and privacy outcomes from a moderated online social therapy for young people with first-episode psychosis. Psychiatric services 2014;65(4):546-550 [FREE Full text] [doi: 10.1176/appi.ps.201300078]
81. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing Messaging to Engage Patients in an Online Suicide Prevention Intervention: Survey Results From Patients With Current Suicidal Ideation. J Med Internet Res 2014 Feb 07;16(2):e42. [doi: 10.2196/jmir.3173]
82. Ghoncheh R, Gould MS, Twisk JW, Kerkhof AJ, Koot HM. Efficacy of Adolescent Suicide Prevention E-Learning Modules for Gatekeepers: A Randomized Controlled Trial. JMIR Ment Health 2016 Jan 29;3(1):e8 [FREE Full text] [doi: 10.2196/mental.4614] [Medline: 26825006]

XSL•FO
RenderX

83. Lamis DA, Underwood M, D'Amore N. Outcomes of a Suicide Prevention Gatekeeper Training Program Among School Personnel. Crisis 2017 Mar;38(2):89-99. [doi: 10.1027/0227-5910/a000414]
84. Smith AR, Silva C, Covington DW, Joiner TE. An assessment of suicide-related knowledge and skills among health professionals. Health Psychology 2014;33(2):110-119. [doi: 10.1037/a0031062]
85. Marshall E, York J, Magruder K, Yeager D, Knapp R, De Santis ML, et al. Implementation of Online Suicide-Specific Training for VA Providers. Acad Psychiatry 2014 Feb 22;38(5):566-574. [doi: 10.1007/s40596-014-0039-5]
86. Krieg C. Can demonstration enhance the effects of an online risk assessment training workshop? Doctoral dissertation, Counseling Psychology. Arizona State University. ASU Library. Digital Repository. 2017. URL: https://repository.asu.edu/items/40196 [accessed 2020-09-17]
87. Rigsbee N. Suicide assessment training in counselor education. Doctoral dissertation. University of New Mexico. UNM Digital Repository. Albuquerque, New Mexico: The University of New Mexico; 2015. URL: https://tinyurl.com/3vzfr7bs [accessed 2020-09-17]
88. Robinson J, Hetrick S, Cox G, Bendall S, Yuen HP, Yung A, et al. Can an Internet-based intervention reduce suicidal ideation, depression and hopelessness among secondary school students: results from a pilot study. Early Interv Psychiatry 2016 Feb;10(1):28-35. [doi: 10.1111/eip.12137] [Medline: 24684946]
89. Skovgaard Larsen JL, Frandsen H, Erlangsen A. MYPLAN – A Mobile Phone Application for Supporting People at Risk of Suicide. Crisis 2016 May;37(3):236-240. [doi: 10.1027/0227-5910/a000371]
90. Rizvi S, Hughes C, Thomas M. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychological Services 2016;13(4):380-388. [doi: 10.1037/ser0000100]
91. Larsen ME, Shand F, Morley K, Batterham PJ, Petrie K, Reda B, et al. A Mobile Text Message Intervention to Reduce Repeat Suicidal Episodes: Design and Development of Reconnecting After a Suicide Attempt (RAFT). JMIR Ment Health 2017 Dec 13;4(4):e56. [doi: 10.2196/mental.7500]
92. AL-Asadi AM, Klein B, Meyer D. Multiple Comorbidities of 21 Psychological Disorders and Relationships With Psychosocial Variables: A Study of the Online Assessment and Diagnostic System Within a Web-Based Population. J Med Internet Res 2015 Feb 26;17(3):e55. [doi: 10.2196/jmir.4143]
93. Hesdorffer DC, French JA, Posner K, DiVentura B, Pollard JR, Sperling MR, et al. Suicidal ideation and behavior screening in intractable focal epilepsy eligible for drug trials. Epilepsia 2013 Feb 28;54(5):879-887. [doi: 10.1111/epi.12128]
94. Best P, Foye U, Taylor B, Hazlett D, Manktelow R. Online interactive suicide support services: quality and accessibility. Mental Health Review Journal 2013 Nov 29;18(4):226-239. [doi: 10.1108/mhrj-03-2013-0009]
95. Flamarique I, Santosh P, Zuddas A, Arango C, Purper-Ouakil D, Hoekstra P, STOP consortium. Development and psychometric properties of the Suicidality: Treatment Occurring in Paediatrics (STOP) Suicidality Assessment Scale (STOP-SAS) in children and adolescents. BMC pediatrics 2016;16(1):213 [FREE Full text] [doi: 10.1186/s12887-016-0751-2]
96. Hirschhorn E. Integrating technology into collaborative suicide risk assessment: comparing electronic and paper-and-pencil versions of the suicide status form. In: Doctoral dissertation. Washington D.C: Catholic University of America; 2017.
97. AL-Asadi AM, Klein B, Meyer D. Comorbidity Structure of Psychological Disorders in the Online e-PASS Data as Predictors of Psychosocial Adjustment Measures: Psychological Distress, Adequate Social Support, Self-Confidence, Quality of Life, and Suicidal Ideation. J Med Internet Res 2014 Oct 28;16(10):e248. [doi: 10.2196/jmir.3591]
98. De Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof AJ. Applying Computer Adaptive Testing to Optimize Online Assessment of Suicidal Behavior: A Simulation Study. J Med Internet Res 2014 Sep 11;16(9):e207. [doi: 10.2196/jmir.3511]
99. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using New and Emerging Technologies to Identify and Respond to Suicidality Among Help-Seeking Young People: A Cross-Sectional Study. J Med Internet Res 2017 Jul 12;19(7):e247. [doi: 10.2196/jmir.7897]
100. Haner D, Pepler D. "Live Chat" Clients at Kids Help Phone: Individual Characteristics and Problem Topics. J Can Acad Child Adolesc Psychiatry 2016;25(3):138-144 [FREE Full text] [Medline: 27924143]
101. Bennett A, Pourmand A, Shokoohi H, Shesser R, Sanchez J, Joyce J. Impacts of Social Networking Sites on Patient Care in the Emergency Department. Telemedicine and e-Health 2014 Jan;20(1):94-96. [doi: 10.1089/tmj.2013.0055]
102. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VWS. FitMindKit : Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interventions 2018 Jun;12:91-99. [doi: 10.1016/j.invent.2017.08.002]
103. Rice S, Robinson J, Bendall S, Hetrick S, Cox G, Bailey E, et al. Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation. J Can Acad Child Adolesc Psychiatry 2016;25(2):80-86 [FREE Full text] [Medline: 27274743]
104. Bresó A, Martínez-Miranda J, Botella C, Baños R, García-Gómez J. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. Expert Systems: International Journal of Knowledge Engineering and Neural Networks 2016;33(4):297-312. [doi: 10.1111/exsy.12151]

XSL•FO
RenderX

105. Madsen T, van Spijker B, Karstoft K, Nordentoft M, Kerkhof AJ. Trajectories of Suicidal Ideation in People Seeking Web-Based Help for Suicidality: Secondary Analysis of a Dutch Randomized Controlled Trial. J Med Internet Res 2016 Jun 30;18(6):e178. [doi: 10.2196/jmir.5904]
106. Hetrick SE, Goodall J, Yuen HP, Davey CG, Parker AG, Robinson J, et al. Comprehensive Online Self-Monitoring to Support Clinicians Manage Risk of Suicide in Youth Depression. Crisis 2017 May;38(3):147-157. [doi: 10.1027/0227-5910/a000422]
107. Hetrick SE, Yuen HP, Bailey E, Cox GR, Templer K, Rice SM, et al. Internet-based cognitive behavioural therapy for young people with suicide-related behaviour (Reframe-IT): a randomised controlled trial. Evid Based Mental Health 2017 Jul 12;20(3):76-82. [doi: 10.1136/eb-2017-102719]
108. Batterham PJ, Christensen H, Mackinnon AJ, Gosling JA, Thorndike FP, Ritterband LM, et al. Trajectories of change and long-term outcomes in a randomised controlled trial of internet-based insomnia treatment to prevent depression. BJPsych open 2018 Jan 02;3(5):228-235. [doi: 10.1192/bjpo.bp.117.005231]
109. Wilksch SM, O'Shea A, Wade TD. Depressive symptoms, alcohol and other drug use, and suicide risk: Prevention and treatment effects from a two-country online eating disorder risk reduction trial. Int J Eat Disord 2018 Dec 23;52(2):132-141. [doi: 10.1002/eat.23005]
110. Maulik PK, Kallakuri S, Devarapalli S. Operational challenges in conducting a community-based technology-enabled mental health services delivery model for rural India: Experiences from the SMART Mental Health Project. Wellcome Open Res 2018 Apr 18;3:43. [doi: 10.12688/wellcomeopenres.14524.1]
111. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? Efficacy of Internet-based guided self-help compared to online psychoeducation for major depressive disorder. Journal of Affective Disorders 2019 Mar;246:695-705. [doi: 10.1016/j.jad.2018.12.065]
112. Richards D, Duffy D, Burke J, Anderson M, Connell S, Timulak L. Supported Internet-Delivered Cognitive Behavior Treatment for Adults with Severe Depressive Symptoms: A Secondary Analysis. JMIR Ment Health 2018 Oct 02;5(4):e10204. [doi: 10.2196/10204]
113. Tewari A, Kallakuri S, Devarapalli S, Jha V, Patel A, Maulik PK. Process evaluation of the systematic medical appraisal, referral and treatment (SMART) mental health project in rural India. BMC Psychiatry 2017 Dec 4;17(1):385. [doi: 10.1186/s12888-017-1525-6]
114. Wilks CR, Lungu A, Ang SY, Matsumiya B, Yin Q, Linehan MM. A randomized controlled trial of an Internet delivered dialectical behavior therapy skills training for suicidal and heavy episodic drinkers. Journal of Affective Disorders 2018 May;232:219-228. [doi: 10.1016/j.jad.2018.02.053]
115. Labelle R, Bibaud-De SA, Leblanc F. Innovating to treat depressionprevent suicide: the IPhone@ PSY ASSISTANCE application. In: Mishara B, Kerkhof AJFM, editors. Suicide Prevention and New Technologies: Evidence-Based Practices. London: Palgrave Macmillan; 2013:166-180.
116. Bush NE, Dobscha SK, Crumpton R, Denneson LM, Hoffman JE, Crain A, et al. A Virtual Hope Box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. Suicide Life Threat Behav 2015 Feb;45(1):1-9. [doi: 10.1111/sltb.12103] [Medline: 24828126]
117. Povey J, Mills PPJR, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of Mental Health Apps for Aboriginal and Torres Strait Islander Australians: A Qualitative Study. J Med Internet Res 2016;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
118. Mackie C, Dunn N, MacLean S, Testa V, Heisel M, Hatcher S. A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. Evid Based Mental Health 2017 Oct 13;20(4):118-122. [doi: 10.1136/eb-2017-102764]
119. Mok K, Jorm AF, Pirkis J. Who Goes Online for Suicide-Related Reasons? Crisis 2016 Mar;37(2):112-120. [doi: 10.1027/0227-5910/a000366]
120. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
121. O'Toole M, Arendt M, Pedersen C. Testing an app-assisted treatment for suicide prevention in a randomized controlled trial: effects on suicide risk and depression. Behavior Therapy 2019;50(2):421-429 [FREE Full text] [doi: 10.1016/j.beth.2018.07.007]
122. Hetrick SE, Dellosa MK, Simmons MB, Phillips L. Development and pilot testing of an online monitoring tool of depression symptoms and side effects for young people being treated for depression. Early Intervention in Psychiatry 2014 Feb 25;9(1):66-69. [doi: 10.1111/eip.12127]
123. Chen RY, Feltes JR, Tzeng WS, Lu ZY, Pan M, Zhao N, et al. Phone-Based Interventions in Adolescent Psychiatry: A Perspective and Proof of Concept Pilot Study With a Focus on Depression and Autism. JMIR Res Protoc 2017 Jun 16;6(6):e114. [doi: 10.2196/resprot.7245]
124. Berrouiguet S, Gravey M, Le Galudec M, Alavi Z, Walter M. Post-acute crisis text messaging outreach for suicide prevention: A pilot study. Psychiatry Research 2014 Jul;217(3):154-157. [doi: 10.1016/j.psychres.2014.02.034]

XSL•FO
RenderX

125. Kodama T, Syouji H, Takaki S, Fujimoto H, Ishikawa S, Fukutake M, et al. Text messaging for psychiatric outpatients: effect on help-seeking and self-harming behaviors. Journal of Psychosocial Nursing and Mental Health Services 2016;54(4):31-37 [FREE Full text] [doi: 10.3928/02793695-20160121-01]
126. Owens C, Charles N. Implementation of a text-messaging intervention for adolescents who self-harm (TeenTEXT): a feasibility study using normalisation process theory. Child and Adolescent Psychiatry and Mental Health 2016;10(1):14 [FREE Full text] [doi: 10.1186/s13034-016-0101-z]
127. Predmore Z, Ramchand R, Ayer L, Kotzias V, Engel C, Ebener P, et al. Expanding Suicide Crisis Services to Text and Chat. Crisis 2017 Jul;38(4):255-260. [doi: 10.1027/0227-5910/a000460]
128. Mokkenstorm JK, Eikelenboom M, Huisman A, Wiebenga J, Gilissen R, Kerkhof AJFM, et al. Evaluation of the 113Online Suicide Prevention Crisis Chat Service: Outcomes, Helper Behaviors and Comparison to Telephone Hotlines. Suicide Life Threat Behav 2016 Aug 19;47(3):282-296. [doi: 10.1111/sltb.12286]
129. Ahuja AK, Biesaga K, Sudak DM, Draper J, Womble A. Suicide on facebook. J Psychiatr Pract 2014 Mar;20(2):141-146. [doi: 10.1097/01.pra.0000445249.38801.d1] [Medline: 24638049]
130. Reyes-Portillo JA, Chin EM, Toso-Salman J, Blake Turner J, Vawdrey D, Mufson L. Using Electronic Health Record Alerts to Increase Safety Planning with Youth At-Risk for Suicide: A Non-randomized Trial. Child Youth Care Forum 2018 Jan 24;47(3):391-402. [doi: 10.1007/s10566-018-9435-4]
131. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Craddock Lee SJ, Rial KV, et al. Developing a Brief Suicide Prevention Intervention and Mobile Phone Application: A Qualitative Report. Journal of Technology in Human Services 2015 Dec 14;33(4):345-357. [doi: 10.1080/15228835.2015.1106384]
132. Tan Z, Liu X, Liu X, Cheng Q, Zhu T. Designing Microblog Direct Messages to Engage Social Media Users With Suicide Ideation: Interview and Survey Study on Weibo. J Med Internet Res 2017 Dec 12;19(12):e381. [doi: 10.2196/jmir.8729]
133. LeMay R. Student learning of violence prevention education concepts: a longitudinal analysis. Doctoral Dissertation. 2015. URL: https://www.semanticscholar.org/paper/Student-Learning-of-Violence-Prevention-Education-A-LeMay/5217862648ffdff4702a835be8497d502f237393#citing-papers [accessed 2020-09-17]
134. de Andrade NNG, Pawson D, Muriello D, Donahue L, Guadagno J. Ethics and artificial intelligenceuicide prevention on Facebook. Philosophy & Technology 2018;31(4):669-684. [doi: 10.1007/s13347-018-0336-0]
135. McKernan LC, Clayton EW, Walsh CG. Protecting Life While Preserving Liberty: Ethical Recommendations for Suicide Prevention With Artificial Intelligence. Front. Psychiatry 2018 Dec 3;9:650. [doi: 10.3389/fpsyt.2018.00650]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
HEAR: Health Education Assessment and Referral
iCBT: internet-based cognitive behavioral therapy
ICT: information and communication technology
PICO: population, intervention, comparison, and outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by N Maltais, E Bailey, K Baudin; comments to author 16.12.20; revised version received 10.02.21; accepted 16.03.21; published 04.05.21.

Please cite as:
Rassy J, Bardon C, Dargis L, Côté LP, Corthésy-Blondin L, Mörch CM, Labelle R
Information and Communication Technology Use in Suicide Prevention: Scoping Review
J Med Internet Res 2021;23(5):e25288
URL: https://www.jmir.org/2021/5/e25288
doi:10.2196/25288
PMID:33820754

©Jessica Rassy, Cécile Bardon, Luc Dargis, Louis-Philippe Côté, Laurent Corthésy-Blondin, Carl-Maria Mörch, Réal Labelle. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2021. 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

Machine Learning and Natural Language Processing in Mental Health: Systematic Review

Aziliz Le Glaz[1], MD; Yannis Haralambous[2], PhD; Deok-Hee Kim-Dufor[1], PhD; Philippe Lenca[2], PhD; Romain Billot[2], PhD; Taylor C Ryan[3], MS; Jonathan Marsh[4], MA; Jordan DeVylder[4], MD, PhD; Michel Walter[1,5], MD, PhD; Sofian Berrouiguet[1,2,5,6], MD; Christophe Lemey[1,2,5], MD

[1]URCI Mental Health Department, Brest Medical University Hospital, Brest, France

[2]IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Brest, France

[3]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[4]Fordham University Graduate School of Social Service, New York, NY, United States

[5]EA 7479 SPURBO, Université de Bretagne Occidentale, Brest, France

[6]LaTIM, INSERM, UMR 1101, Brest, France

Corresponding Author:
Christophe Lemey, MD
URCI Mental Health Department
Brest Medical University Hospital
Route de Ploudalmézeau
Brest, 29200
France
Phone: 33 619211032
Email: christophe.lemey@chu-brest.fr

Abstract

Background: Machine learning systems are part of the field of artificial intelligence that automatically learn models from data to make better decisions. Natural language processing (NLP), by using corpora and learning approaches, provides good performance in statistical tasks, such as text classification or sentiment mining.

Objective: The primary aim of this systematic review was to summarize and characterize, in methodological and technical terms, studies that used machine learning and NLP techniques for mental health. The secondary aim was to consider the potential use of these methods in mental health clinical practice

Methods: This systematic review follows the PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) guidelines and is registered with PROSPERO (Prospective Register of Systematic Reviews; number CRD42019107376). The search was conducted using 4 medical databases (PubMed, Scopus, ScienceDirect, and PsycINFO) with the following keywords: machine learning, data mining, psychiatry, mental health, and mental disorder. The exclusion criteria were as follows: languages other than English, anonymization process, case studies, conference papers, and reviews. No limitations on publication dates were imposed.

Results: A total of 327 articles were identified, of which 269 (82.3%) were excluded and 58 (17.7%) were included in the review. The results were organized through a qualitative perspective. Although studies had heterogeneous topics and methods, some themes emerged. Population studies could be grouped into 3 categories: patients included in medical databases, patients who came to the emergency room, and social media users. The main objectives were to extract symptoms, classify severity of illness, compare therapy effectiveness, provide psychopathological clues, and challenge the current nosography. Medical records and social media were the 2 major data sources. With regard to the methods used, preprocessing used the standard methods of NLP and unique identifier extraction dedicated to medical texts. Efficient classifiers were preferred rather than transparent functioning classifiers. Python was the most frequently used platform.

Conclusions: Machine learning and NLP models have been highly topical issues in medicine in recent years and may be considered a new paradigm in medical research. However, these processes tend to confirm clinical hypotheses rather than developing entirely new information, and only one major category of the population (ie, social media users) is an imprecise cohort. Moreover, some language-specific features can improve the performance of NLP methods, and their extension to other languages should be more closely investigated. However, machine learning and NLP techniques provide useful information from

XSL•FO
RenderX

unexplored data (ie, patients' daily habits that are usually inaccessible to care providers). Before considering It as an additional tool of mental health care, ethical issues remain and should be discussed in a timely manner. Machine learning and NLP methods may offer multiple perspectives in mental health research but should also be considered as tools to support clinical practice.

(J Med Internet Res 2021;23(5):e15708) doi:10.2196/15708

KEYWORDS

machine learning; natural language processing; artificial intelligence; data mining; mental health; psychiatry

Introduction

Machine Learning

Machine learning (ML) systems automatically learn models from data to make better decisions. As such, they are part of a major subfield of artificial intelligence (AI). There are 3 main approaches to learning from data: supervised, unsupervised, and reinforcement learning. In supervised learning, a target attribute is predicted, and ML algorithms infer a model from labeled input data (ie, a training data set that provides examples described by predictive attributes and values for the target attribute). The goal is to make target predictions on new data to obtain good generalization performance. In contrast, there is no target attribute in unsupervised learning, and thus no labeled data. Unsupervised learning consists of inferring a model to describe hidden patterns from unlabeled data. Under circumstances in which labeled data acquisition proves to be difficult, (eg, costly), semisupervised ML methods can use both labeled and unlabeled data for learning. The third main category of ML is reinforcement learning, in which the ML model uses feedback that acts as a reward or punishment to maximize its performance.

ML is limited to certain capacities. For one, it relies on collections of data that may be incomplete, noisy, or subject to systematic bias, all of which can lead to erroneous predictions. In addition, ML algorithms may introduce bias. Interesting questions to be addressed in ML are discussed in an article by Domingos [1]. However, when carefully conducted, ML can have great utility.

AI and ML have many applications, many of which are encountered in daily life. Supervised ML, for example, is widely used for spam filtering (ie, classifying incoming email as spam or not spam) [2]. It is also used to classify credit applicants based on their probabilities of default [3]. Unsupervised ML, such as algorithm clustering, is able to group customers with similar characteristics and their likelihood to purchase. This is widely used by banks for market segmentation [4]. Finally, automatic document clustering that organizes similar documents into classes (for purposes of improving information retrieval, for example) is gaining importance due to the increasing number of documents on the internet [5].

The application of ML in health is also of concern. Indeed, ML is widely used in critical disease models in cardiology, neurology, and diabetes research [6] to automatically identify heart disease risk factors [7], to classify primary progressive aphasia subtypes [8], and for the characterization and diagnosis of cognitive impairments [9], diabetes, and cardiovascular disorders [10-17].

ML is also challenging the traditional epidemiologic approach of evidence-based medicine owing to its high processing speed and ability to handle large volumes of data with heterogeneous variables (electronic health records, administrative data sets, wearable sensors, genomic and proteomic databanks, and social media) [18]. In fact, AI and ML have huge potential to build inferences and find patterns in vast volumes of patient histories, medical images, epidemiological statistics, and other particulars such as natural language data. For example, they can help doctors improve their diagnoses, forecast disease outbreaks, and customize treatments [19,20], provide better patient care [21], and predict the splicing activity of individual exons and chromatin marks from DNA sequences [22]. From a mental health perspective, the prevention of suicidal risk has recently been substantially studied [23-26].

Indeed, mental health care is also benefiting from the advancements in ML [27-29]. Classical ML with only mixed data (observations described by a mixture of numerical and categorical variables) is widely used, but language-based deficits are common symptoms of depression, bipolar disorder, autism spectrum disorder (ASD), personality disorder, and schizophrenia [30]. This implies that computational linguistics could have a great role in forming new insights into individuals' mental health and emotions.

Language in both spoken and written forms plays an important role in ML mental health applications. It is therefore essential to understand what natural language processing (NLP) is before discussing the joint applications of ML and NLP in mental health.

NLP

NLP is a subdiscipline of computer science that emerged in the 1960s. In 1967, the first published book on the subject, *Introduction to Computational Linguistics* [31], clearly considers language from a symbolic point of view: it describes techniques such as syntax parsing using dependency trees or Chomsky transformational grammars and statistical methods (word counting) are only hinted at. At that time, computing resources were sparse and had to be carefully managed; hence, a whole chapter of the book is dedicated to the storage of grammars in memory. The situation changed in the 1990s when personal computers became largely available and increasingly powerful. A new approach to NLP based on statistical methods emerged. The book by Manning and Schütze, *Foundations of Statistical Natural Language Process* [32], is a landmark of this evolution [32]. The 3 main sections of the book are dedicated to (1) methods at the word level (collocations, *n*-grams, and word sense disambiguation), (2) methods at the sentence level (morphosyntactic parsing using Markov models, and

XSL•FO
RenderX

probabilistic context-free grammars), and (3) clustering, classification, and information retrieval. Probabilistic context-free grammars are a typical example of the evolution of NLP methods: the symbolic approach by Chomsky—or at least a simplified version—is endowed with probabilities attached to productions, and the ambiguity of natural language is reflected in the coexistence of several syntax trees with different probabilities.

During the same period, symbolic methods evolved as well. The 1990s witnessed the emergence of the World Wide Web, the Semantic Web, and ontology engineering. First, the 2 research directions seemed contradictory. Knowledge representation was aimed at structuring knowledge in an exhaustively precise symbolic manner, whereas the statistical viewpoint considered language in the same way as physics considers natural phenomena: by analyzing them through various heteroclitic methods, identifying general laws by numerical indicators, and proving them using statistical methods. An example illustrating the latter is the distributional semantic hypothesis (originally stated in the paper by Harris titled, Distributional structure [33]) asserting that "Words occurring in the same contexts will tend to have related meanings." According to this hypothesis, one does not need to identify the precise meaning of a word, as a symbolic method would require, but simply to find the word's cooccurrences in a corpus and consider these as semantics of the word. A very popular method called latent semantic analysis (LSA) is based on the following: the matrix of occurrences of words in documents (contexts) is reduced so that the dimensions of the new matrix represent aggregates of words and aggregates of documents where each dimension is not interpretable per se, but when words or documents are represented as vectors in this new *latent* system of coordinates, the scalar product of vectors can be used as a semantic relatedness measure [34]. LSA is also an example of a typical ML method, with a learning phase (when the frequencies of words in the corpus are counted and the word or document matrix is reduced) to perform a specific task (evaluating the similarity between documents).

Since the 2000s and 2010s, a new evolution has occurred in NLP with the emergence of convolutional, recurrent, and recursive neural networks (NNs) [35]. By using large corpora and sophisticated learning approaches, these methods provide good performance in tasks of statistical nature, such as text classification or sentiment mining. In the past 3 years, they have been much more frequently used for learning higher syntactic or semantic structures (syntax graphs or concept mining, respectively).

In the future, hybrid methods may be used more frequently, which combine symbolic and statistical approaches. The presence of ML methods in NLP systems is a trend that will undoubtedly remain integral to contemporary methods through the foreseeable future.

Applications of ML and NLP to Mental Health

Applications of ML and NLP to mental health can be classified according to the following axes:

- The corpus: as one of the topics is NLP, the corpus necessarily has a textual component. The most common corpora are records or reports (electronic health records [EHRs], Psychological Evaluation Reports, and Coroner Reports), social media (Reddit, Twitter, etc), or transcribed patient interviews.
- Corpus processing: depending on the nature of the corpus, one can either extract medical terms and match them with unified medical language system (UMLS) concept unique identifiers (CUIs) or process blocks of text in natural language and perform specific searches (eg, to detect terms related to suicide).
- Classification methods: many ML techniques are used, such as decision trees, support vector machines, conditional random fields, random forests, and NNs.
- Goal: the goal is usually to validate a hypothesis or to study the behavior of a given population of patients.

Corpora can be of a very large volume. For instance, Sinnenberg et al [36] published a systematic review about Twitter as a tool for health research that included 137 different studies and analyzed over 5 billion tweets using ML; Castro et al [37] have processed 4.2 million EHRs spanning a period of over 20 years. Corpora can also be small, as demonstrated in the study conducted by Carson et al [38], who treated 73 respondents' unstructured clinical notes, or in the study by Bedi et al [39], in which only 34 participants' 1-hour-long narrative interviews were considered. Sometimes, corpora are created specifically for a project. For example, in a study by Roy et al [40], volunteers had written 182 abusive texts, annotated by researchers and abuse victims, and these texts were then analyzed and provided a model for detecting abusive texts.

Extraction of the UMLS CUIs is mainly applied to EHRs because the latter are semistructured and constitute a special document type. The specificities of this document type are reflected in its structure, the syntax of text, and, most importantly, the vocabulary used. The extraction of medical terms is achieved through information extraction algorithms and matching these terms with UMLS CUIs is performed through knowledge representation methods. Once these concepts have been extracted from an EHR, the latter is represented by the former and concepts become features used for classification.

On corpora other than EHRs, rather than extracting the UMLS CUIs, more general NLP methods are applied to textual data to obtain features that are then classified by ML algorithms. These NLP methods are often frequency counts of words or *n*-grams in a specific set, which can be manually curated or obtained out of a corpus. In other cases, methods such as LSA or latent Dirichlet allocation (LDA) are used for topic detection. The initial set of words can be explicit. For example, Doan et al [41] collected tweets containing the hashtags #relax and #stressed and classified them by theme and location. In other cases, calculations are performed at a higher level and words involved in the process are not explicitly known. For example, Luo et al [42] attempted to characterize autism by analyzing textual descriptions of closely related individuals written by patients or members of a control group. Nevertheless, most NLP applications in mental health rely on words (using the bag-of-words method, that is, ignoring word order and keeping

XSL•FO
RenderX

only their frequencies). Some take word order into account in a limited way (by using *n*-grams, ie, contiguous sequences of words of length *n*), but very few take syntax into account by the use of dependency trees [18,43,44]. With respect to their applications, it should be noted that ML and NLP tools are invaluable in alleviating data issues such as data overflow in modern medicine. Forsting et al [45] acknowledge that ML and NLP techniques can be useful for optimism bias (eg, the difference between a person's expectation and the actual outcome or the concept that a clinician may think that his or her patient's problem falls solely into a specific discipline in which the physician works) because the machine has a generalist approach unlike the specialist clinician. Within the last two decades, these techniques have emerged in mental health, following the success of social media to act as an informative source of data [46].

In addition, NLP is essential in psychiatry because language-based deficits are common symptoms of depression, behavioral disorder, ASD, personality disorder, and schizophrenia [30]. It can provide insight into individuals' mental health and emotions, their use of narrative, subjective, and structured speech styles, and their lifestyle, specifically their educational level, socioeconomic status, living conditions, and cultural background [47], all of which are routine in mental status examinations.

Using ML in general and NLP methods in particular, one can create semiautomated systems (operating under human supervision) aiming to improve the specificity of diagnosis, knowledge of psychophysiology, speed of diagnosis, and more accurate estimations of disease severity [48]. Through analyses of Twitter posts, O'Dea et al [49] identified the importance of creating real-time campaigns to increase help-seeking behaviors and reduce the stigma attached to mental health. Moreover, automated programs can be more cost-effective and time-efficient than their traditional counterparts. Ly et al [50] proposed using interventions based on an automated self-help system as a way to make mental health promotion tools more widely accessible. In addition, Lucas et al [51] demonstrated through a clinical trial that when people believed they were interacting with a computer rather than an actual clinician, they reported less fear of self-disclosure, reported reduced impression management behaviors, experienced more ease in expressing the severity of their emotions, and were rated by observers as more willing to disclose. However, these findings may not be generalizable, as they were potentially biased by their sample selections and/or system design itself.

Although ML and NLP provide new tools and strategies for psychiatric research and practice [52], it should be kept in mind that their use frequently raises ethical and legal concerns over consent to personal data use and data anonymization. Similarly, studies using AI for predictive analyses are challenging the balance between beneficence and respect for patients' autonomy. McKernan et al [53] suggest that efforts be made to communicate AI methods to obtain free and informed consent from patients. Moreover, prospective studies should be conducted to evaluate the use of AI tools [53].

The primary aim of this systematic review is to summarize and characterize studies that used ML and NLP techniques for mental health in methodological and technical terms. Hence, the secondary aim is to consider the potential use of these methods in mental health clinical practice, such as the contributions that they may offer in areas of diagnosis and prognosis, the establishment of risk factors, impacts of psychotherapy, treatment adherence, and side effects.

Methods

This systematic review is grounded in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) guidelines [54]. Searches were carried out as specified by the standard protocol for PROSPERO (Prospective Register of Systematic Reviews; registration number CRD42019107376).

Literature Search Strategy

A systematic, computerized literature search was conducted using 4 databases: PubMed (via MEDLINE), Scopus, ScienceDirect, and PsycINFO. Each database was explored from August 21, 2018, through February 1, 2020, with no publication date limit. The search was carried out using the following keywords: "natural language processing" AND "machine learning" AND ("psychiatry" OR "mental health" OR "mental disorder"). The same search was performed on the element (data mining) instead of (machine learning). When the full text was not available, the abstract was used to extract the necessary information to avoid selection bias. Case studies, conference papers, and reviews were excluded.

Study Selection and Eligibility Criteria

After removing duplicates, 2 collaborators independently screened all titles and abstracts that were relevant to this systematic review. A third reviewer was consulted when disagreement arose between the first 2 collaborators. The process is depicted in Multimedia Appendix 1. Only studies available in English were selected. We deliberately excluded studies about the anonymization process to focus on the articles investigating the clinical use of ML and NLP in psychiatry (eg, contribution to diagnosis, prognosis, establishment of risk factors, impact of psychotherapy, treatment adherence, and side effect). No limitations on publication dates were imposed. A total of 58 articles were included in the review.

Included Studies

All studies were thoroughly screened, and their main ideas are summarized in individual tables (Multimedia Appendix 2 [37-41,43,47,48,55-104]). These tables provide information on qualitative and quantitative features: authors, year of publication, precise topic of mental health (eg, autism, psychotic spectrum disorder, etc), population characteristics, and types and volume of recorded data. The second part of these tables summarizes the objectives, methods, and results.

XSL•FO
RenderX

Results

Study Selection

The database search resulted in 222 studies identified using the (machine learning) keyword and 105 studies using the (data mining) keyword. After merging them, 238 unique studies were considered for review, based on the title and abstract. A total of 84 papers were excluded because (1) they were not about psychiatry or mental health (52 cases), (2) they were not written in English (1 case), and (3) the keywords (machine learning), (natural language processing), or (data mining) did not appear in the title or abstract (8 cases). As a second filter, 33 studies about data anonymization were excluded. Furthermore, 7 studies were excluded because ML or NLP were not their main subject but were only quoted as background information. In addition, 96 papers were excluded because they were reviews, case studies, or conference papers. Finally, 58 articles were included in this review.

Topics and Population

Topics are heterogeneous. The most frequently mentioned topics are depression and suicide with 17 studies [38,55,57,60-62,77-79,82,83,87,88,91,92,99,104]. Other psychiatric diagnoses were addiction to alcohol or illicit drugs (6 cases) [43,65,66,75,84,86]; posttraumatic stress disorder (PTSD; 3 cases) [47,63,64]; neurodevelopmental disorders (3 cases) [42,58,93]; psychotic spectrum disorders, including schizophrenia (3 cases) [39,95,100]; anxiety (2 cases) [41,98]; personality disorder (1 case) [85]; eating disorders (2 cases) [89,96]; and bipolar disorder (2 cases) [37,102]. A total of 3 studies were on violence and cyber harassment [40,80,94]. Treatment issues such as adherence or misuse are also depicted (6 cases) [56,72,74,81,101,103]. Only 1 study on mechanical restraints [90] and 1 on cognitive troubles [97] were found. A total of 8 studies were transnosographic [59,67-71,73,76]: 6 met the CEGS N-GRID 2016 Center of Excellence in Genomic Science Neuropsychiatric-Genome-Scale and Research Domain Criteria (RDoC) Individualized Domains 2016 Shared Task in Clinical NLP criteria, which will be developed further in our results.

In total, 3 distinct categories of population were found:

1. Patients whose EHRs were available in science-based research databases such as the Partners HealthCare electronic medical record (EMR), a collection of data from patients at Massachusetts General Hospital and Brigham and Women's Hospital [55,56]. These records extended beyond psychiatric records and included other medical records as well.
2. Patients seen in emergency or psychiatry departments who had additional clinical characteristics in their records (eg, clinical observation, laboratory tests, diagnostic and therapeutic interventions, typed specialists' notes).
3. Social media networks (Facebook, Twitter, and Instagram): The authors of these studies have selected specific hashtags such as #stress or #depression and have screened a multitude of public messages using a streaming platform.

Objectives

In total, 5 main categories of objectives were found: to extract clinical symptoms, to classify severity of illnesses, to compare different therapies, to provide psychopathological clues in mental health, and to challenge the current nosography.

The principal objectives of these studies were to extract and record clinical symptoms, establish a diagnosis, or monitor changes over time. A total of 2 studies targeted automated epidemiological monitoring: Metzger et al [57] provided a method of detecting suicide attempts from EHRs and Leroy et al [58] achieved automatic extraction of criteria for ASD from EHRs with an accuracy of 76%. The latter study stated that an increasing prevalence of given symptoms (nonverbal behavior, social and emotional reciprocity, and adherence to routine disabilities) occurred from 2000 through 2010. Data extraction was also used for diagnosis: He et al [47] diagnosed PTSD with an accuracy of 82% after analyzing free texts written by trauma survivors.

In addition to extraction, an important aim was to measure the severity of psychiatric disorders in psychological evaluation record corpora. Goodwin et al [59] classified symptoms of patients with psychosis into 4 different levels of severity (absent, mild, moderate, and severe) using statistical analyses. Fernandes et al [60] studied EHRs from a cohort of individuals with a history of suicide attempts and a cohort of individuals with a history of suicidal ideation only. Their algorithm of detecting suicidal ideation or suicide attempts had a sensitivity of 98.2% and a positive predictive value of 82.8% [57]. Other studies found that ML and NLP techniques performed well, although they were not necessarily better than a practitioner's ability to predict the clinical risk of suicide in their patients [61,62]; thus, the authors proposed statistical NLP approaches to be used in collaboration with clinical practice.

ML and NLP methods are also used to measure and compare the effectiveness of different types of psychotherapy [63,64]. Tanana et al [43] investigated 2 statistical NLP techniques to code motivational interviewing sessions. Motivational interviewing is a psychotherapy method used for substance use disorders and other behavioral problems to strengthen personal motivation for change [105]. Motivational interviews can be manually coded to assess therapy adherence and gather feedback for subsequent sessions. The authors found that the discrete sentence feature model (a sentence classifier based on *n*-gram models) had accuracy similar to the manual coding of therapeutic sessions. Maguen et al [63] used statistical NLP techniques to distinguish evidence-based psychotherapy, including cognitive processing therapy and prolonged exposure notes from unstructured psychotherapy notes for a population of veterans with PTSD. They found that almost 20% of veterans observed an improvement in their symptoms after one or more sessions of evidence-based psychotherapy.

Another objective was to provide psychopathological clues for understanding mental health disorders by analyzing language features. This objective sometimes involves the processing of previously unexplored data, such as chat groups or social networks. The following are some examples of studies that pursue this objective: Baggott et al [65] found that MDMA

XSL•FO
RenderX

(3,4-méthylènedioxy-N-méthylamphétamine; Ecstasy) altered individuals' speech patterns more frequently than the placebo and led to an increase in both positive and negative social and sexual language use (others, public, camaraderie, and outgoing). Chary et al [66] analyzed posts on Lycaeum, a popular web forum known for being one of the most frequently cited platforms with respect to drug use. They discovered new combinations of drugs that were not mentioned in the medical literature. Luo et al [42] differentiated the social interactions between adults with ASD and healthy adults. They confirmed the hypothesis regarding differences in language and social interactions in adults with ASD: typical participants had more connected semantic links than the ASD group and the words with the largest number of connections were different between the 2 groups. Doan et al [41] noticed that American Twitter users are more likely to express their source of stress on Twitter than in their day-to-day experiences. The main causes of stress that emerged from the Twitter data were education, work, and social relationships. They also found that individuals' expressions of stress and relaxation differed based on the city of residence (Los Angeles, New York, San Diego, and San Francisco). Moreover, Mowery et al [106] revealed that less than 2% of the tweet corpus (a corpus of 9300 annotated tweets containing depression-related keywords) included more than one depression-related reference, suggesting that there may be different forms of expression when it comes to depression.

Finally, AI in mental health research challenges the current practice and nosography. In 2010, Insel et al [107] initiated a project called the RDoC, a research framework for mental health disorders that aims to constitute an alternative to the DSM (Diagnostic and Statistical Manual of Mental Disorders). The former includes data on genetics and neuroscience in its classification of mental health disorders, whereas the latter is solely based on clinical data [107]. The RDoC is a matrix in which the columns and rows represent constructs (genes, molecules, cells, circuits, physiology, behaviors, self-reports, and paradigms) and subconstructs of each of the following 6 domains: negative valence, positive valence, cognitive systems, systems for social processes, arousal or regulatory systems, and sensorimotor systems. Pro-RDoC practitioners argue that DSM syndromes have significant limitations when used as phenotypes for identifying biomarkers and specific genetic variants associated with mental illness [108]. A concrete application of this new system used statistical NLP methods to create a phenotypic homogenous cohort that allowed a better comparison [109]. In 2016, the CEGS N-GRID (Centers of Excellence in Genomic Science Neuropsychiatric-Genome-Scale and RDoC Individualized Domains) proposed 3 challenging tasks using NLP methods: (1) data anonymization, (2) predicting symptom severity in the positive valence domain from neuropsychiatric clinical records, and (3) novel data use case (eg, predicting the presence of common mental conditions in patients) [67]. This research on NLP and ML processing identified 6 articles [59,67-71] that met these challenge tasks, although only 1 study dealt with task 3 [67]. As mentioned earlier, studies on anonymization were excluded; thus, the RDoC framework links the neuro-biological basis of mental processes with phenotypical manifestations [110]. The CEGS N-GRID shared task provided usable data for investigating ML and NLP techniques, which could lead to new psychiatric nosology.

Type of Data Used

As can be seen in Table 1 (in which no hapaxes are displayed), the most frequent corpus type is that of EHRs (to which EMRs can be added). EHRs (and EMRs) are convenient data sources because of their heterogeneity: they combine structured, semistructured, and free data, and they often use a significantly controlled language containing medical terms that allow the extraction of CUIs (Methods Section). The second most frequent sources of data are clinical notes and clinical records, which share the convenient properties of EHRs or EMRs, but are not standardized in the same way.

Table 1. Corpus type.

Characteristics	Values
EHRs[a]	22.9508
ClinNotes	16.3934
ClinRecords	11.4754
Interviews	8.1967
Tweets	8.1967
Questionnaires	6.5574
Reddit	6.5574
Web	4.918
EMRs[b]	3.2787

[a]EHR: electronic health record.

[b]EMR: electronic medical record.

The data described earlier share an important property: the corpora are generated by practitioners and therefore can be used for medical term extraction with satisfactory results.

A different category of data is generated by the patients. This category can be divided into 2 subcategories: data generated with the help of practitioners (eg, interviews and questionnaires)

XSL•FO
RenderX

and data freely generated by patients on social media (tweets, posts on Reddit, web blogs, etc).

Interviews and (textual parts of) questionnaires are technically free text but practitioners still have some amount of control over the content, and the environment in which the data are collected influences the degree of informality of texts. For these reasons, traditional NLP methods can be applied to them with satisfactory results.

Data collected from social media, because of their high degree of informality, loose spelling and syntax, and use of abbreviations and emojis, can only be superficially processed by standard NLP methods. Typical examples are in studies by Doan et al [41] and Jackson et al [73], in which tweets were selected because they contained the hashtags #stress and #relax and their words were used in a bag-of-words without any further linguistic treatment [41] or tweets were selected based on the presence of terms denoting opioids [72]. Although the authors lemmatized tweet contents, the main feature of tweets taken into account was their geographical origin.

Methodology

Two phases of NLP projects were distinguished: (1) preprocessing, which consists of analyzing the data to obtain numeric or categorical features, and (2) classification.

Preprocessing

Table 2 (in which no hapaxes are displayed) represents the frequency of use of various preprocessing methods that can be of different natures. Some methods apply to words or word groups: *lemma (*lemmatization, ie, replacing a word by a base form such as the singular for nouns or the infinitive for verbs), *POS* (part of speech, ie, attaching to a word a label denoting its grammatical function), *cTAKES* or *CUIs* (mapping a word or a noun phrase to concept in an ontology, such as the UMLS, and therefore unambiguously defining its semantics), *tf-idf* (attaching to a word or a term a value representing its significance in characterizing a given document or class it belongs to), *embedding* (representing a word by a vector in a high-dimensional space), *named-entity recognition* (deciding whether a given word or noun phrase is a named entity), *LIWC* (Linguistic Inquiry and Word Count, a commercial tool advertised as being “based on solid science” providing various “social and psychological insights” of words). Other methods combine words into higher structures: *n-grams* (considering an *n*-gram, ie, a sequence of *n* subsequent words, as an entity and measuring the frequencies of these entities). Finally, other methods are applied to entire sentences, paragraphs, or documents: *SentiAna* (analyzing sentiments or emotions), *LDA* and *LSA* (calculating sets of topics, detecting the significance of each topic for a given document, and providing representative words for each topic). The most frequent preprocessing methods are the standard methods of NLP (lemmatization, part-of-speech tagging, *n*-grams, and tf-idf), and methods specific to medical texts such as CUI extraction (keywords cTAKES and CUIs in Table 2). The embedding method is related almost exclusively to NNs and therefore is relatively recent. Finally, the tail of the graph in Table 2 contains methods applied primarily to free texts such as topic detection, named-entity recognition, sentiment or emotion analysis.

Table 2. Preprocessing methods.

Characteristics	Values
lemma	16.3043
POS[a]	10.8696
cTAKES[b]	10.8696
ngrams	9.7826
tfidf	7.6087
embedding	6.5217
CUIs[c]	5.4348
LDA[d]	5.4348
SentiAna	5.4348
LIWC[e]	4.3478
NER[f]	4.3478
LSA[g]	3.2609

[a]POS: part of speech.

[b]cTAKES: clinical Text Analysis and Knowledge Extraction System.

[c]CUI: concept unique identifier.

[d]LDA: latent Dirichlet allocation.

[e]LIWC: Linguistic Inquiry and Word Count.

[f]NER: named-entity recognition

[g]LSA: latent semantic analysis.

XSL•FO
RenderX

Classification

Once the classification phase is reached, linguistic data are entirely converted into numeric data, and therefore, the choice of classifier depends on factors other than corpus type. Some of these factors include (1) the volume of data, (2) the type of classification (supervised vs unsupervised), (3) the explicability level, and (4) the platform used. In Table 3 (where hapaxes are not displayed), we have shown decision tree, association rules, and C4.5 (also a decision tree algorithm) that are *transparent* methods, that is, the user can follow the classification process in a step-by-step manner and understand the reason a given individual belongs to a particular class. They are not the most frequent classifiers, probably because explicability is not a major concern of most studies. Instead, the most frequently used classifiers such as support vector machine (SVM), LogiR (logistic regression), RF (random forest), and LinR (linear regression) are solid, fast legacy classifiers with small parameter sets and good performance. In the middle of Table 3 are NNs that belong to the deep learning tendency of ML: they are opposite to DT/AR/C4.5 when it comes to explicability and they rely heavily on certain parameters (type and geometry of NN, number of layers, size of layers, optimizer, learning rate, loss function, etc). The causes of the relatively low frequency of NNs in publications may be (1) the fact that they have been implemented in user-friendly frameworks (such as Theano or Keras) only recently, (2) the necessity to fine-tune a large number of parameters, and (3) the relatively high requirements in terms of memory, central processing unit, and graphical processing unit. This is likely to change in the near future.

Table 3. Classifier type.

Characteristics	Values
SVM[a]	22.6804
LogiR[b]	16.4948
RF[c]	11.3402
DT[d]	6.1856
NB[e]	6.1856
NN[f]	6.1856
LinR[g]	5.1546
K-Means	3.0928
AR[h]	2.0619
C4.5	2.0619

[a]SVM: support vector machine.
[b]LogiR: logistic regression.
[c]RF: random forest.
[d]DT: decision tree.
[e]NB: Naive Bayes.
[f]NN: neural network.
[g]LinR: linear regression.
[h]AR: association rules.

Platforms

As can be seen in Table 4 (hapaxes are not represented), the 2 most common platforms are Python and R. Python is a *universal* programming language, in the sense that it is not specific to a given domain: more than 120,000 packages allow the user to perform specialized tasks in any possible field. Furthermore, it is open-source and high-quality documentation abounds. R is also an open-source programming language and compiler, but contrary to Python, it is oriented toward statistics. Although many classifiers have been implemented efficiently both in Python and R, the domain of NLP is better represented in Python, in credit to packages such as NLTK (Natural Language ToolKit), spaCy, and Stanza. The third bar, titled *Unknown*, represents publications that do not mention the platform used. The fourth bar indicates the General Architecture for Text Engineering General Health platform, an open-source Java application that provides an environment for processing textual data in a user-friendly manner. The *Apache* bar gathers different tools distributed by the Apache Software Foundation. Stata is a commercial statistics software from College Station, Texas, first released in 1985. Weka is an open-source programming environment for ML.

Figure 1 shows the use of platforms in chronological order. The use of Python and R started after 2015, while Stata, Weka, and Apache were already in use in 2011.

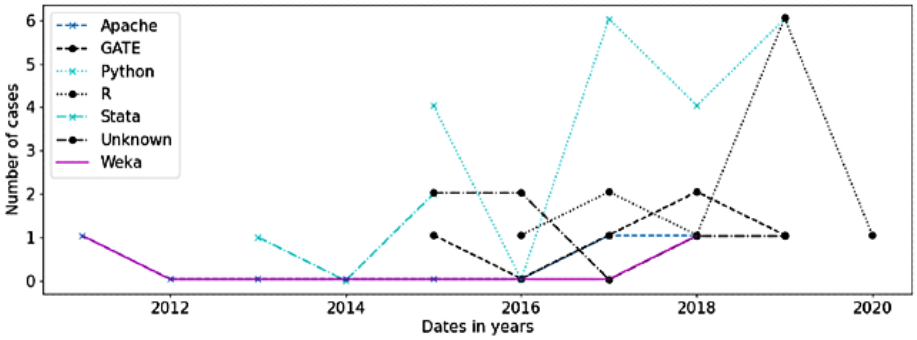
XSL•FO
RenderX

Table 4. Platforms.

Characteristics	Values
Python	34.4828
R	18.9655
Unknown	10.3448
GATE[a]	8.6207
Apache	5.1724
Stata	5.1724
Weka	3.4483

[a]GATE: General Architecture for Text Engineering General Health.

Figure 1. Platforms usage.

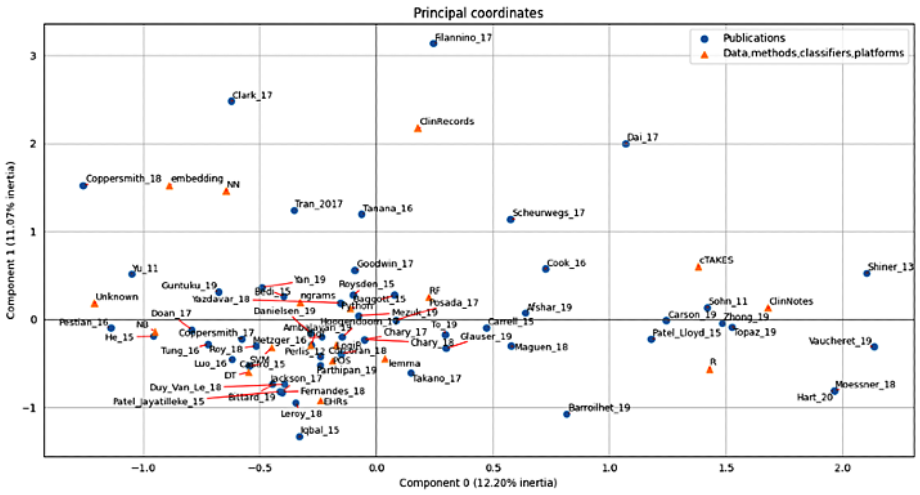


Correspondence Analysis of Data, Methods, Classifiers, Platforms, and Publications

The correspondence analysis is a dimension reduction technique that maps the data into a factorial space where each dimension is a combination of the initial variables. Figure 2 represents the principal coordinates of the publications and the various entities considered in their study.

On the right, a cluster of publications is surrounded by data type *ClinNotes*, method *cTAKES*, and platform *R*. In the upper left quadrant, some publications gather with method *embedding* and classifier *NN*. Toward the left of the diagram and close to the horizontal axis, publications with an *unknown* platform using the *NB* classifier are present along with a big cluster whose center includes *tf-idf*, *LogiR*, *SVM*, *Python*, and *n-grams*: the legacy methods, most used classifiers, and the most used platform.

Figure 2. Correspondence analysis.



With regard to publications, *Filannino_17* is an obvious outlier because it has no method, classifier, or platform and because it describes a task and how this task has been treated by others. *Clark_17* is at the extreme upper left, as it uses NNs and k-means (the latter is not displayed because only entities appearing at least 5 times are included). *Coppersmith_18* also uses embeddings and NNs, whereas *Tran_17* (which is closer to the central cluster) uses both NNs and SVMs. On the right side, *Shiner_13* and *Vaucheret_19* use clinical notes and R, whereas *Hart_20* and *Moessner_18* use R and methods that

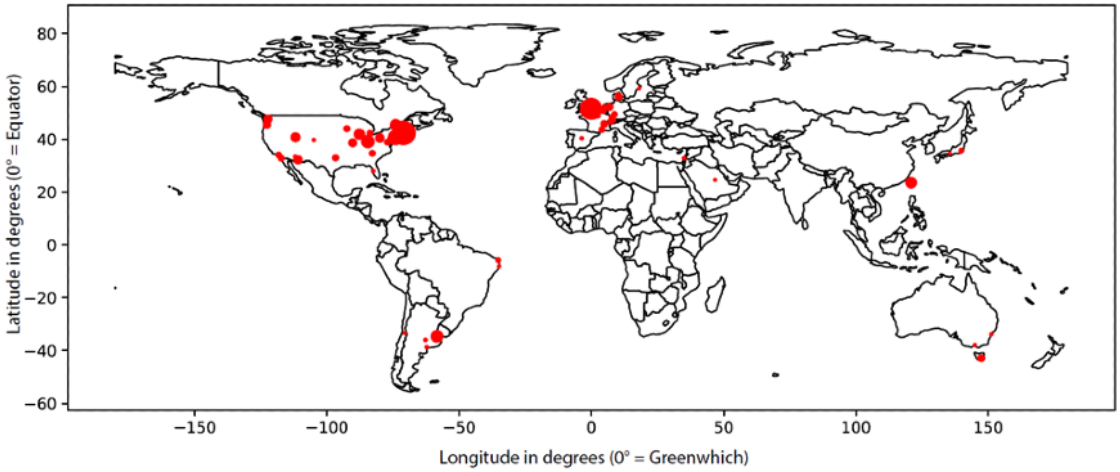
XSL•FO
RenderX

have not been taken into account in the calculation. In the bottom left, *Iqbal_15* uses EHRs in the General Architecture for Text Engineering General Health (which is not displayed). At the extreme left and close to the horizontal axis, *Pestian_16* and *Yu_11* use an unknown platform.

Geographical Distribution of Authors

In the map in Figure 3, the diameter of the red marks is proportional to a score calculated as follows: we added 1 unit for the geographical origin of the affiliation of each author of each paper. The cities with scores greater than 10 were Boston (54), London (44), New York (21), Cincinnati (15), Buenos Aires (13), Cambridge, Massachusetts (12), San Francisco (11), and Taiwan (11).

Figure 3. Geographical distribution of authors.



Citations and Cocitations

Figure 4 represents the citations of the papers in our list by other papers on the same list. The size of the nodes of a paper is proportional to the number of papers citing it. The colors of the nodes and edges represent communities. Each community has a central node: Perlis et al [55] are cited in 7 other papers, Jackson et al [73] are cited in 4 other papers, Carrell et al [74] and Afshar et al [75] are cited in 2 other papers, and Bedi et al [39] are cited in 2 other papers. In total, 22 papers are singletons: they are neither cited nor cite any other paper in our list.

Although mutual citations show influences between papers in our list, we can also measure the number of cocitations (ie, common references between 2 papers in the list). In Figure 5, the edges between papers indicate that they have at least 3 common references. The edge width is proportional to the number of references.

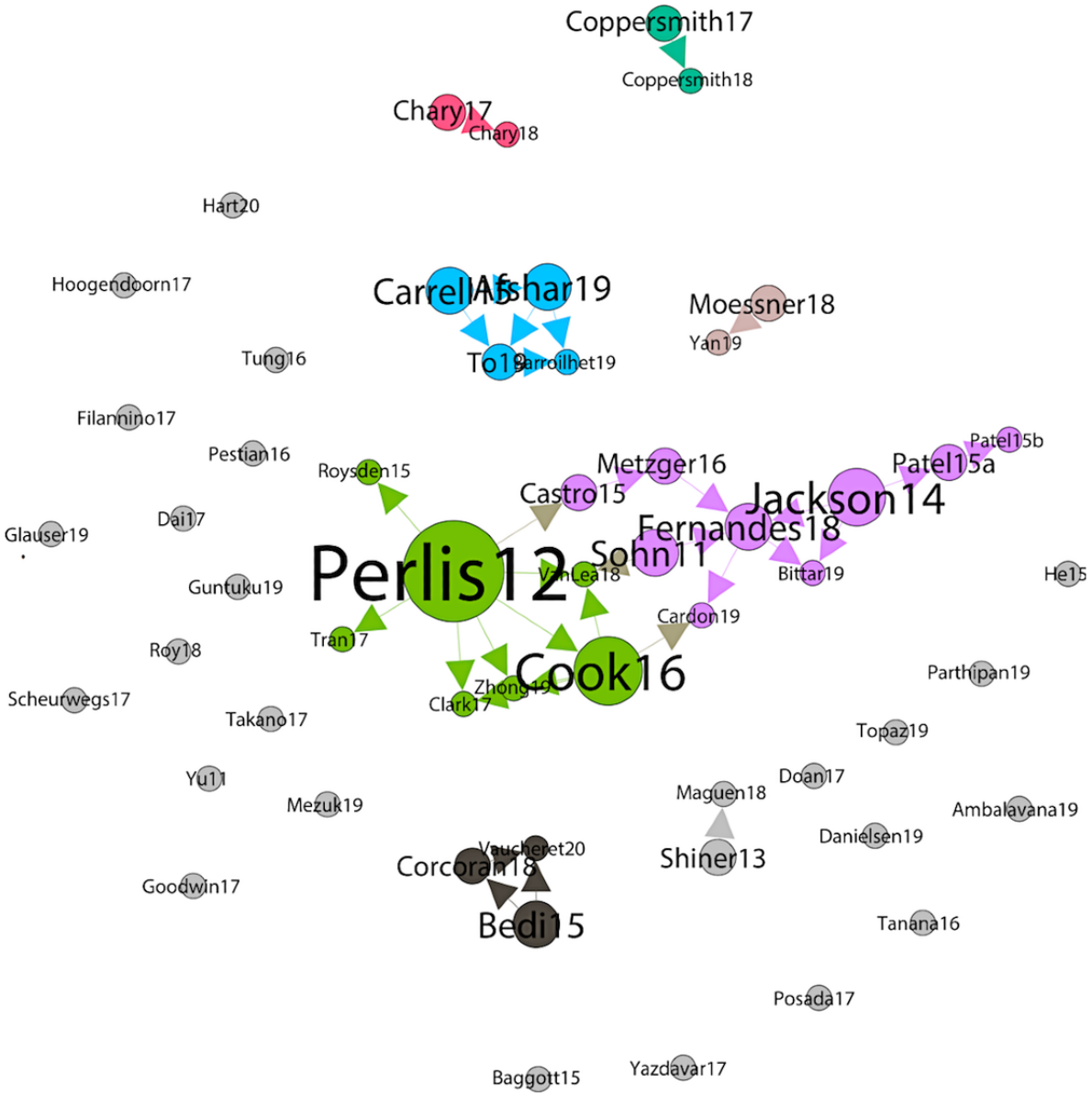
The edge of the greatest width is the one between the papers by Coppersmith et al [76] and Coppersmith et al [77], which is normal—the 2 papers share the same first author, have been released within less than a year, and have 26 common references.

The second case, in descending order of edge width, is between Shiner et al [64] and Maguen et al [63]. This is also normal—the first author of the former is also the last author of the latter and the latter is presented as an extension of the former: “In this study, our goal was to extend Shiner and colleagues’ work by applying automated coding to a large national pool of mental health treatment notes in order to identify the use of cognitive processing therapy and prolonged exposure.” The 2 papers share 14 common references.

The size of nodes in the graph is proportional to the degree. Zhong et al [78] have the highest degree: this paper has more than three common references with as many as eight other papers, in fact, with 8 references. The color of the nodes and edges corresponds to the connected components. There are 19 singleton nodes that share ≤2 references with every other paper of the list.

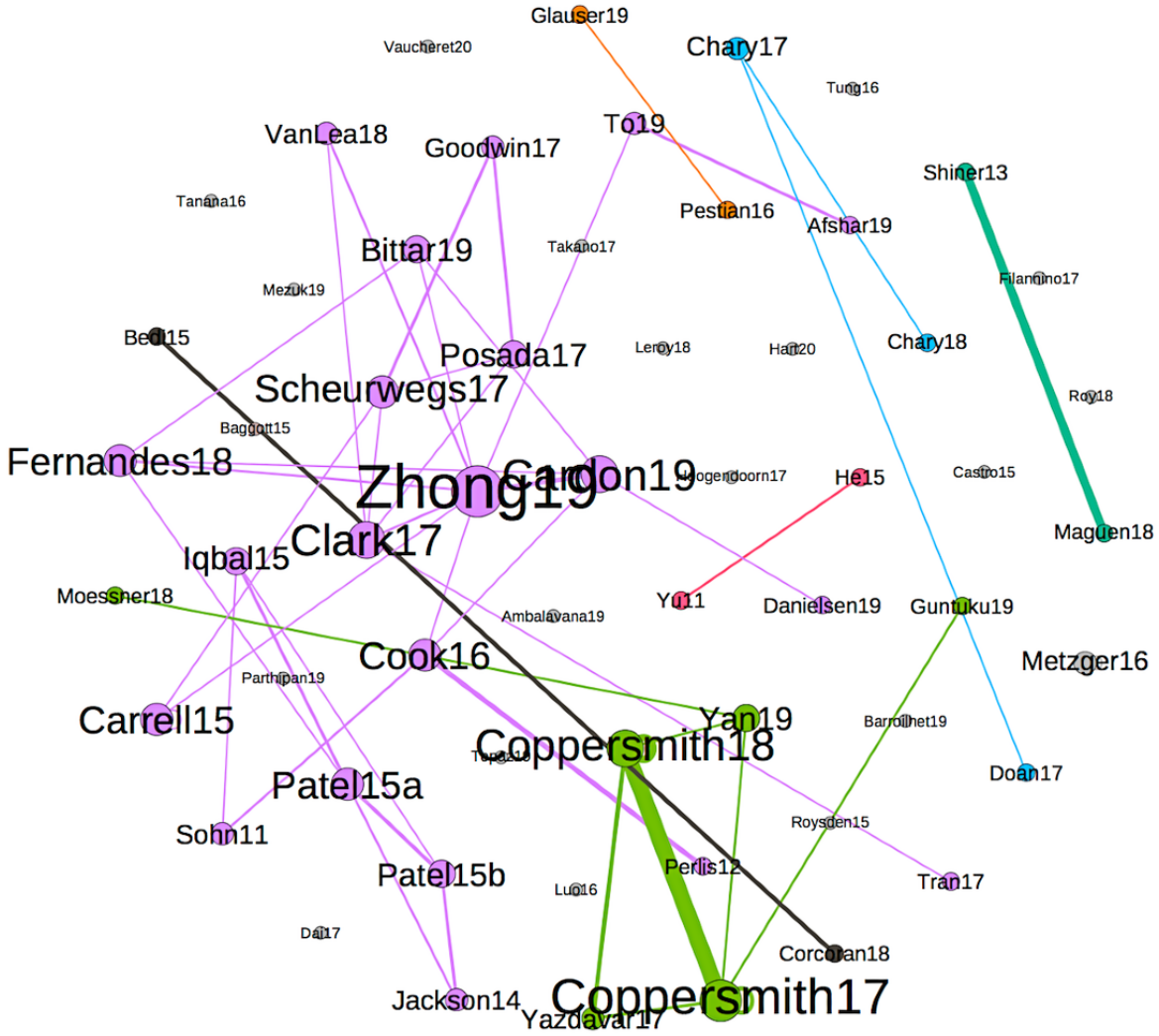
XSL•FO
RenderX

Figure 4. Graph of cocitations.



XSL•FO
RenderX

Figure 5. Graph of cocitations.



Discussion

Strengths and Limitations of the Review

This study reviews ML and NLP models in the field of mental health, which has been a highly topical issue in recent years. The methodology was elaborated to screen a maximum number of specific medical studies by expanding the research to 4 medical databases (PubMed, Scopus, ScienceDirect, and PsycINFO). Furthermore, the characterization of the selected studies has been done very precisely in a qualitative manner to simultaneously depict the populations, methods, data sources, and technical aspects.

The primary limitation of this study is the lack of quantitative comparisons between the selected studies. It is indeed not feasible to compare highly heterogeneous studies that do not share common research patterns. In addition, the selected works were not scored on their risk of bias. Despite this shortcoming, their limitations and strengths are outlined in the individual tables in Multimedia Appendix 2.

Methodological and Technical Limitations of the Selected Studies

ML and NLP methods may be considered as a new paradigm in medical research in which it becomes practical to analyze every possible, even unexpected, and innovative parameter of a topic to discern new clinical patterns. This new paradigm involves reconsidering the standard methodology, which consists of formulating a sound hypothesis, defining objectives, and collecting results to either uphold or reject the hypothesis. However, in practice, the selected studies tend to confirm clinical hypotheses based on fundamental clinical intuitions, namely language abnormalities in adults with ASD [42].

Other methodological limitations and potential bias sources have been noted. As stated in the Results section, one of the 3 main population categories is *social network or chat users* [40,41,66,77,79], whose members are predominantly young. Owing to this, Coppersmith et al [76,77] cautioned that these results may not be generalizable to other populations [77,106]. In addition, when Chary et al [66] focused on *Lycaeum users* and Coppersmith et al [76] mentioned *participants from a company*, the lack of precise information on the participants of a cohort was obvious. An exception to this is the group of OurDataHelps.org users [77] who volunteered to participate in

XSL•FO
RenderX

scientific research and filled out a questionnaire to provide information about themselves. Even when participants volunteer to provide personal information, there is a high likelihood that personality bias plays a role, especially in studies on suicide and depression.

Similarly, studies rarely consider cultural or ethnic differences within a sample [80]. For example, in a study on violent behavior, researchers should acknowledge that *spanking children for discipline purposes* is considered inappropriate in some cultures but appropriate in others. In some cases, language-specific features can improve the performance of NLP methods. For example, in the case of Takano et al [62], the distribution of morphemes is used to distinguish between specific and nonspecific memories in the Autobiographical Memory Test. As shown in the paper, among the most important distinctive factors are grammatical particles that are specific to the Japanese language, such as た/だ (past tense), ない (negation), は (topic marker), and で (place or method). In languages with different structures, the same method may be less efficient and other indicators may need to be investigated.

Is There an Advantage in Using ML and NLP for Mental Health Clinical Practice?

The hallmark ML principle is to simultaneously analyze large quantities of data; however, this sometimes leads researchers to the implicit assumption that *the more data they input, the more accurate will be the results*. ML and NL allow the analysis of large amounts of data and the comparison of broad groups and patients. For example, Roysden et al [56] screened administrative data and EHRs from a population of 12,759 patients; Maguen et al [63] compared over 8,168,330 clinical notes collected over 15 years; and Yazdavar et al [79] analyzed posts authored by 4000 Twitter users. At the same time, even though thousands of papers have been published using medical data, very few have made meaningful contributions to clinical practices [111].

Twitter and other social networks, with almost 3 billion users globally, have become significant sources of information for medical use and research [112]. Moreover, the analysis of social media–based platforms can generate valuable details about people's mental health and social or professional interactions. The alteration of daily habits is one of the core criteria for the diagnosis of a mental health disorder (in general, criterion B of DSM-5). A recent study by Fagherazzi and Ravaud [113] illustrates the idea that AI can be implemented in the so-called *digitosome* (data generated online and by digital technologies) that constitutes a powerful agent for detecting new digital markers and risk factors in medicine. By analyzing a global cohort of more than 85,000 tweets per week authored by people with diabetes, they were able to discuss different illness-related stress patterns of patients with type 1 or type 2 diabetes. By analyzing tweets, Mowery et al [106] found that there may be alternative ways in which people express depression. These findings indicate that there may be new ways for people to express mental illness.

From this perspective, different expressions of psychological distress (whether people are addressing health care professionals, relatives, or digital friend networks) could be accessible and useful to care providers. ML and NLP may be valuable in psychiatry for identifying people with clinical risks for depression, suicide attempts, anxiety, or even psychosis based on digital data or clinical notes.

Ethical Reflections

AI in psychiatry and more broadly in medicine raises ethical issues and requires prudence in its application. As mentioned earlier, ML and NLP techniques have valuable advantages in psychiatry for analyzing large amounts of data with high diagnostic and prognostic validity. These tools, which have been groundbreaking in medicine and psychiatry, should receive more attention for their promising results with regard to clinical practice and medical research. In addition, recent studies suggest that people are becoming more comfortable when speaking with a machine compared with a clinician: Lucas et al [51] state that in a clinical trial, people who (believed they) were interacting with a computer disclosed information more openly than people who thought that an individual was controlling the computer. Perhaps the machine is viewed as being more objective than a human and therefore reduces the fear of judgment from a practitioner. The introduction of a computer in medical practice as a new type of clinician leads to a profound change in the physician-patient relationship and promotes the idea of having a new clinical model involving a third party. The relationship is crucial to psychiatric clinical practice, and the use of data processing should be discussed. Sassolas [114] questioned this *technological psychiatry* as a practice that is likely to avoid what he called the "psychic privacy proximity." Technological psychiatry could generate an operative encounter whose unique purpose is to normalize the patient's symptoms and reduce the fear of disclosure.

In addition to improved relationships, the application of ML and NLP in psychiatry should be done with special precautions to avoid clinical abuse. This review includes 2 studies about the prediction of psychosis in patients at high risk of this disease. One even introduced a model of ML+NLP that had a 100% accuracy in predicting psychosis among the latter patient sample [39], which was better than a simple clinical evaluation. Nevertheless, these results should be treated with caution because of the small sample size and the lack of detail on the statistical techniques used. The risk of overfitting needs to be considered. Although further research should be continued to improve technical issues, ethics should be taken into account. Martinez-Martin et al [115] questioned whether it is ethical to use prognostic estimates from ML to treat psychosis, as it is not known whether variables are present in the local context (such as differences in psychiatric practice and social support) that would affect the model's validity. Moreover, when programming an ML algorithm, investigators can choose to strengthen the criteria they esteem to be more relevant, such as clinical criteria instead of socioeconomic factors. This could result in loss of opportunity for some patients when the automated machine analysis gives the illusion of greater objectivity. These adjustments should be done to respect the principle of equity.

In the case of predicting psychosis, the study involved only patients who consented to both psychiatric care and the completion of interviews. This was not the case in studies on

XSL•FO
RenderX

suicide prevention, where researchers tracked information on patients by using social media. This could be considered a violation of confidentiality. Should information from social media be used to identify symptoms? Applying AI in this context raises significant ethical concerns, particularly in balancing beneficence and respecting confidentiality [53]. ML and NLP can help identify people at clinical risk for depression or suicidal ideation, who most likely do not have access to mental health providers and/or a primary care doctor [61]; however, this reduces confidentiality protection and can lead to increased vulnerability in certain populations [21]. To obtain informed consent from patients and protect their privacy, McKernan et al [53] proposed some recommendations: patients should be informed that (1) algorithms can be imperfect or wrong; (2) algorithm data should be considered highly sensitive or confidential; (3) algorithm data might recommend actions that are not immediately apparent; and (4) algorithms might prompt unnecessary intervention from the provider. Therefore, psychiatrists should be trained in ML and NLP techniques and be able to explain to patients their main characteristics and why they may require certain recommendations. This last point underlines the need for an explainable AI that goes further than black box methods.

Finally, ML and NLP should not lead to disempowerment of psychiatrists or replace the clinician-patient pair. On the contrary, the combination of ML with NLP should be considered as a *tool* to support clinical practice and medical research.

Conclusions

In the past decade, the use of ML and NLP has become increasingly widespread in medicine and more specifically in psychiatry. Hence, this review aimed to summarize and characterize studies that used ML and NLP techniques for mental health in methodological and technical terms. The secondary aim was to consider the potential use of these methods in mental health clinical practice (eg, contribution to diagnosis, prognosis, establishment of risk factors, impact of psychotherapy, treatment adherence, and side effects).

Although the selected studies were heterogeneous in terms of topics and mental disorders, common features were found in terms of population categories (patients included in medical databases, patients presenting to the emergency room, and social media network users) and objectives (ie, symptom extraction, severity classification, comparison of therapies, findings of psychopathological clues, and challenges to the current nosography). The type-of-data-used analysis identified 2 major corpora: data collected by care providers (EHR, clinical notes, or EMR) and data from social media. Finally, the method analysis indicates that the authors privileged certain techniques. The standard methods of NLP (such as lemmatization, POS tagging, or n-grams) are most frequently used for preprocessing, in addition to CUI extraction dedicated to medical texts. The classification analysis specifies that classifiers with good performance (SVM, LogIR, and RF) are preferred to those with *transparent* functioning. The use of the universal programming language platforms such as Python and R is verified; Python turned out to be the most frequently and recently used. The correspondence analysis of data, methods, classifiers, platforms, and publications reveals a cluster of publications associating clinical notes data with cTAKES methods and the R-Python platform.

ML and NLP methods may sometimes be impressive with their huge amount of data screening and the multiple perspectives they offer. This has led some authors to consider it to be a new paradigm in mental health research. However, these processes tend to confirm clinical hypotheses rather than developing new information, and some results should be treated with caution (eg, results from social media users' cohorts or the impact of language-specific features on NLP methods performance). On the contrary, ML and NLP techniques provide information from unexplored data and on patients' daily habits that are usually inaccessible to care providers. It may be considered as an additional tool in every step of mental health care: diagnosis, prognosis, treatment efficacy, and monitoring. In this regard, ethical issues, such as predicting psychiatric troubles or implications in the physician-patient relationship, remain and should be discussed in a timely manner. Therefore, ML and NLP methods may offer multiple perspectives in mental health research, but they should be considered as a tool to support clinical practice.

Acknowledgments

This work has been supported by the French government's "Investissements d'Avenir" programme, which is managed by the Agence Nationale de la Recherche (ANR), under the reference ANR-18-RHUS-0014.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred reporting items for systematic reviews (PRISMA) flow diagram.
[PNG File , 160 KB - jmir_v23i5e15708_app1.png]

Multimedia Appendix 2
Table summarizing the selected studies.
[PDF File (Adobe PDF File), 338 KB - jmir_v23i5e15708_app2.pdf]

XSL•FO
RenderX

References

1. Domingos P. A few useful things to know about machine learning. Commun ACM 2012 Oct;55(10):78-87. [doi: 10.1145/2347736.2347755]
2. Dada EG, Bassi JS, Chiroma H, Abdulhamid SM, Adetunmbi AO, Ajibuwa OE. Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 2019 Jun;5(6):e01802 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e01802] [Medline: 31211254]
3. Lessmann S, Baesens B, Seow H, Thomas LC. Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur J Oper Res 2015 Nov;247(1):124-136. [doi: 10.1016/j.ejor.2015.05.030]
4. Danuta Z, Jan M. Clustering Algorithms for Bank Customer Segmentation. In: 5th International Conference on Intelligent Systems Design and Applications. 2005 Presented at: ISDA'05; June 26-27, 2005; Warsaw. [doi: 10.1109/isda.2005.33]
5. Afreen M, Badugu S. Document clustering using different unsupervised learning approaches: A survey. In: Advances in Decision Sciences, Image Processing, Security and Computer Vision. Cham, UK: Springer; 2000.
6. 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]
7. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, et al. An automatic system to identify heart disease risk factors in clinical texts over time. J Biomed Inform 2015 Dec;58 Suppl:S158-S163 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.002] [Medline: 26362344]
8. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 2014 Jun;55:43-60. [doi: 10.1016/j.cortex.2012.12.006] [Medline: 23332818]
9. Oliva J, Serrano JI, del Castillo MD, Iglesias A. A methodology for the characterization and diagnosis of cognitive impairments-application to specific language impairment. Artif Intell Med 2014 Jun;61(2):89-96. [doi: 10.1016/j.artmed.2014.04.002] [Medline: 24813116]
10. Mahabub A. A robust voting approach for diabetes prediction using traditional machine learning techniques. SN Appl Sci 2019 Nov 25;1(12):1. [doi: 10.1007/s42452-019-1759-7]
11. Sonark P, JayaMalini. K. Diabetes Prediction Using Different Machine Learning Approaches. In: 3rd International Conference on Computing Methodologies and Communication. 2019 Presented at: ICMMC'19; May 6-9, 2019; Erode, India. [doi: 10.1109/ICCMC.2019.8819841]
12. Sharma M, Singh G, Singh R. An advanced conceptual diagnostic healthcare framework for diabetes and cardiovascular disorders. ICST Transactions on Scalable Information Systems 2018 Jun 19;5(18):154828. [doi: 10.4108/eai.19-6-2018.154828]
13. Cruz J, Wishart D. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform 2017 Oct 04;2:117693510600200-117693510600217 [FREE Full text] [doi: 10.1177/117693510600200030]
14. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
15. Gauthier E, Brisson L, Lenca P, Ragusa S. Breast Cancer Risk Score: a Data Mining Approach to Improve Readability. In: The International Conference on Data Mining. 2011 Presented at: CDM'11; June 23-27, 2011; Las Vegas, United States p. 15-21.
16. Gauthier E, Brisson L, Lenca P, Clavel-Chapelon F, Ragusa S. Challenges to Building a Platform for a Breast Cancer Risk Score. In: Sixth International Conference on Research Challenges in Information Science. 2012 May Presented at: RCIS'12; June 1-7, 2012; Valencia, Spain.
17. Dhahri H, Al Maghayreh E, Mahmood A, Elkilani W, Faisal Nagi M. Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. J Healthc Eng 2019;2019(2):4253641-4253643 [FREE Full text] [doi: 10.1155/2019/4253641] [Medline: 31814951]
18. Scott IA. Machine learning and evidence-based medicine. Ann Intern Med 2018 Jul 3;169(1):44-46. [doi: 10.7326/M18-0115] [Medline: 29710098]
19. Bughin J. Artificial intelligence, the next digital frontier? McKinsey Global Institute 2017 Jun:80.
20. David C, Sastry C, Martha I, Saif R, Jordan V. Ma- Chine Learning and Therapeutics 2.0: Avoiding Hype, Realizing Potential. McKinsey & Commpany. URL: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/machine-learning-and-therapeutics-2-0-avoiding-hype-realizing-potential?cid=eml-app [accessed 2018-12-17]
21. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. J Am Med Assoc 2016 Feb 9;315(6):551-552. [doi: 10.1001/jama.2015.18421] [Medline: 26864406]
22. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol 2016 Jul 29;12(7):878 [FREE Full text] [doi: 10.15252/msb.20156651] [Medline: 27474269]
23. Lopez-Castroman J, Nogue E, Guillaume S, Picot MC, Courtet P. Clustering suicide attempters: impulsive-ambivalent, well-planned, or frequent. J Clin Psychiatry 2016 Jun;77(6):e711-e718. [doi: 10.4088/JCP.15m09882] [Medline: 27035768]

XSL•FO
RenderX

24. Oh J, Yun K, Hwang JH, Chae JH. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front Psychiatry 2017;8:192 [FREE Full text] [doi: 10.3389/fpsyt.2017.00192] [Medline: 29038651]
25. Berrouiguet S, Billot R, Larsen ME, Lopez-Castroman J, Jaussent I, Walter M, et al. An approach for data mining of electronic health record data for suicide risk management: database analysis for clinical decision support. JMIR Ment Health 2019 May 7;6(5):e9766 [FREE Full text] [doi: 10.2196/mental.9766] [Medline: 31066693]
26. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019 Jun 1;76(6):642-651. [doi: 10.1001/jamapsychiatry.2019.0174] [Medline: 30865249]
27. Luxton DD. An introduction to artificial intelligence in behavioral and mental health care. In: Luxton DD, editor. Artificial Intelligence in Behavioral and Mental Health Care. San Diego CA: Academic Press; 2016:1-26.
28. Kaur P, Sharma M. Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis. J Med Syst 2019 May 28;43(7):204. [doi: 10.1007/s10916-019-1341-2] [Medline: 31139933]
29. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
30. Cohen AS, Mitchell KR, Elvevåg B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. Schizophr Res 2014 Nov;159(2-3):533-538 [FREE Full text] [doi: 10.1016/j.schres.2014.09.013] [Medline: 25261880]
31. Hays DG. Introduction to Computational Linguistics, Mathematical Linguistics and Automatic Language Processing. Cambridge: American Elsevier Publishing Co; 1967.
32. Manning CD, Schutze H. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press Ltd; 1999.
33. Harris ZS. Distributional Structure. WORD 2015 Dec 4;10(2-3):146-162. [doi: 10.1080/00437956.1954.11659520]
34. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse Processes 1998 Jan;25(2-3):259-284. [doi: 10.1080/01638539809545028]
35. Chollet F. Deep Learning with Python. Shelter Island: Manning Publications; 2017.
36. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
37. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer G, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
38. Carson NJ. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. PLoS One 2019 Feb 19;14(2):01-14. [doi: 10.1371/journal.pone.0211116] [Medline: 30779800]
39. Bedi G, Carrillo F, Cecchi GA, Fernandez Slezak D, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015 Aug 26;1:15030. [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]
40. Roy T, McClendon J, Hodges L. Analyzing Abusive Text Messages to Detect Digital Dating Abuse. In: IEEE International Conference on Healthcare Informatics. 2018 Presented at: IEEE'18; March 2-3, 2018; New-York. [doi: 10.1109/ICHI.2018.00039]
41. Doan S, Ritchart A, Perry N, Chaparro JD, Conway M. How do you #relax when you're #stressed? A content analysis and infodemiology study of stress-related tweets. JMIR Public Health Surveill 2017 Jun 13;3(2):e35 [FREE Full text] [doi: 10.2196/publichealth.5939] [Medline: 28611016]
42. Luo SX, Shinall JA, Peterson BS, Gerber AJ. Semantic mapping reveals distinct patterns in descriptions of social relations in adults with autism spectrum disorder. Autism Res 2016 Aug;9(8):846-853. [doi: 10.1002/aur.1581] [Medline: 26613541]
43. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Srikumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. J Subst Abuse Treat 2016 Jun;65:43-50 [FREE Full text] [doi: 10.1016/j.jsat.2016.01.006] [Medline: 26944234]
44. Haralambous Y, Lemey C, Lenca P, Billot R, Kim-Dufor DH. Using NLP Methods to Predict Psychiatric Comorbidities. In: the RaPID-3: Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive, psychiatric and/or developmental impairments Workshop. 2020 Presented at: RAPID'20; May 2-6, 2020; Marseille.
45. Forsting M. Machine Learning Will Change Medicine. J Nucl Med 2017 Mar;58(3):357-358 [FREE Full text] [doi: 10.2967/jnumed.117.190397] [Medline: 28153954]
46. Sarasohn-Kahn J. The wisdom of patients: health care meets online social media. Technical report, California HealthCare Foundation 2008 [FREE Full text]
47. He Q, Veldkamp BP, Glas CA, de Vries T. Automated Assessment of Patients' Self-Narratives for Posttraumatic Stress Disorder Screening Using Natural Language Processing and Text Mining. Assessment 2017 Mar;24(2):157-172. [doi: 10.1177/1073191115602551] [Medline: 26358713]

XSL•FO
RenderX

48. Alanazi HO, Abdullah AH, Qureshi KN. A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. J Med Syst 2017 Apr;41(4):69. [doi: 10.1007/s10916-017-0715-6] [Medline: 28285459]
49. O'Dea B, Achilles MR, Larsen ME, Batterham PJ, Calear AL, Christensen H. The rate of reply and nature of responses to suicide-related posts on Twitter. Internet Interv 2018 Sep;13:105-107 [FREE Full text] [doi: 10.1016/j.invent.2018.07.004] [Medline: 30206524]
50. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot rct using mixed methods. Internet Interv 2017 Oct 10;10:39-46. [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
51. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
52. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun;19(6):e228. [doi: 10.2196/jmir.7215] [Medline: 28663166]
53. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving libertythical recommendations for suicide prevention with artificial intelligence. Front Psychiatry 2018;9:650. [doi: 10.3389/fpsyt.2018.00650] [Medline: 30559686]
54. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic reviewmeta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 2;350:g7647. [doi: 10.1136/bmj.g7647] [Medline: 25555855]
55. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. Psychological Medicine 2012 Jan;42(1):41-50. [doi: 10.1017/S0033291711000997] [Medline: 21682950]
56. Roysden N, Wright A. Predicting health care utilization after behavioral health referral using natural language processing and machine learning. 2015 Presented at: AMIA Annu Symp Proc; 2015; San Francisco p. 2063-2072 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765610/pdf/2246862.pdf
57. Metzger MH, Tvardik N, Gicquel Q, Bouvry C, Poulet E, Potinet-Pagliaroli V. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. Int J Methods Psychiatr Res 2017 Jun;26(2):e1522. [doi: 10.1002/mpr.1522] [Medline: 27634457]
58. Leroy G, Gu Y, Pettygrove S, Galindo MK, Arora A, Kurzius-Spencer M. Automated extraction of diagnostic criteria from electronic health records for autism spectrum disorders: Development, evaluation, and application. J Med Internet Res 2018 Nov 7;20(11):e10497. [doi: 10.2196/10497] [Medline: 30404767]
59. Goodwin TR, Maldonado R, Harabagiu SM. Automatic recognition of symptom severity from psychiatric evaluation records. J Biomed Inform 2017 Nov:71-74. [doi: 10.1016/j.jbi.2017.05.020] [Medline: 28576748]
60. Fernandes AC, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep 2018 May 09;8(1):7426. [doi: 10.1038/s41598-018-25773-2]
61. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput Math Methods Med 2016:8708434. [doi: 10.1155/2016/8708434] [Medline: 27752278]
62. Pestian JP, Grupp-Phelan J, Bretonnel Cohen K, Meyers G, Richey LA, Matykiewicz P, et al. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide Life Threat Behav 2016 Apr;46(2):154-159. [doi: 10.1111/sltb.12180] [Medline: 26252868]
63. Maguen S, Madden E, Patterson OV, DuVall SL, Goldstein LA, Burkman K, et al. Measuring use of evidence based psychotherapy for post-traumatic stress disorder in a large national healthcare system. Adm Policy Ment Health 2018 Jul;45(4):519-529. [doi: 10.1007/s10488-018-0850-5] [Medline: 29450781]
64. Shiner B, D'Avolio LW, Nguyen TM, Zayed MH, Young-Xu Y, Desai RA, et al. Measuring use of evidence based psychotherapy for posttraumatic stress disorder. Adm Policy Ment Health 2013 Jul;40(4):311-8. [doi: 10.1007/s10488-012-0421-0] [Medline: 22535469]
65. Baggott MJ, Kirkpatrick MG, Bedi G, de Wit H. Intimate insight: MDMA changes how people talk about significant others. J Psychopharmacol 2015 Jun;29(6):669-677. [doi: 10.1177/0269881115581962]
66. Chary M, Yi D, Manini AF. Candyflipping and Other Combinations: Identifying Drug-Drug Combinations from an Online Forum. Front Psychiatry 2018;9:135 [FREE Full text] [doi: 10.3389/fpsyt.2018.00135] [Medline: 29760666]
67. Tran T, Kavuluru R. Predicting mental conditions based on «history of present illness» in psychiatric notes with deep neural networks. J Biomed Inform 2017 Nov;75:138-148. [doi: 10.1016/j.jbi.2017.06.010] [Medline: 28606869]
68. Scheurwegs E, Sushil M, Tulkens S, Daelemans W, Luyckx K. Counting trees in random forests: Predicting symptom severity in psychiatric intake reports. J Biomed Inform 2017 Nov;75:112-119. [doi: 10.1016/j.jbi.2017.06.007] [Medline: 28602906]
69. Posada JD, Barda AJ, Shi L, Xue D, Ruiz V, Kuan PH, et al. Predictive modeling for classification of positive valence system symptom severity from initial psychiatric evaluation records. J Biomed Inform 2017;75:94-104. [doi: 10.1016/j.jbi.2017.05.019] [Medline: 28571784]

XSL•FO
RenderX

70. Clark C, Wellner B, Davis R, Aberdeen J, Hirschman L. Automatic classification of RDoC positive valence severity with a neural network. Journal of Biomedical Informatics 2017 Nov;75:120-128. [doi: 10.1016/j.jbi.2017.07.005] [Medline: 28694118]
71. Filannino M, Stubbs A, Uzuner O. Symptom severity prediction from neuropsychiatric clinical records: Overview of 2016 CEGS N-GRID shared tasks Track 2. Journal of Biomedical Informatics 2017 Nov;75:62-70. [doi: 10.1016/j.jbi.2017.04.017] [Medline: 28455151]
72. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol 2017;13(4):278-286. [doi: 10.1007/s13181-017-0625-5] [Medline: 28831738]
73. Jackson RG, Ball M, Patel R, Hayes RD, Dobson RJB, Stewart R. Texthunter- a user friendly tool for extracting generic concepts from free text in clinical research. 2014 Nov Presented at: AMIA Annu Symp Proc; 2014; Washington DC p. 729-738 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420012/pdf/1985470.pdf
74. Carrell DS, Cronkite D, Palmer RE, Saunders K, Gross DE, Masters ET, et al. Using natural language processing to identify problem usage of prescription opioids. Int J Med Inform 2015;84(12):1057-1064. [doi: 10.1016/j.ijmedinf.2015.09.002] [Medline: 26456569]
75. Afshar M, Philips A, Karnik N, Mueller J, To D, Gonzalez R, et al. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. Journal of the American Medical Informatics Association 2019;26(3):254-261. [doi: 10.1093/jamia/ocy166 PMCID: PMC6657384] [Medline: 30602031]
76. Coppersmith G, Hilland C, Frieder O, Leary R. Scalable mental health analysis in the clinical whitespace via natural language processing. 2017 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2017; Orlando p. 393-396. [doi: 10.1109/BHI.2017.7897288]
77. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:-11. [doi: 10.1177/1178222618792860] [Medline: 30158822]
78. Zhong QY, Mittal LP. Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. European Journal of Epidemiology 2019 Feb;34(2):153-162. [doi: 10.1007/s10654-018-0470-0] [Medline: 30535584]
79. Yazdavar AH. Semi-supervised approach to monitoring clinical depressive symptoms in social media. 2017 Jul Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pages 1191-1198; 2017; Los Alamitos CA. [doi: 10.1145/3110025.3123028]
80. Dai HJ, Su ECY, Uddin M, Jonnagaddala J. Exploring associations of clinical and social parameters with violent behaviors among psychiatric patients. Journal of Biomedical Informatics 2017;75:149-159. [doi: 10.1016/j.jbi.2017.08.009]
81. Hart KL. What do patients learn about psychotropic medications on the web? A natural language processing study. Journal of affective disorders 2020 Jan 1;260:366-371. [doi: 10.1016/j.jad.2019.09.043] [Medline: 31539672]
82. Glauser T, Santel D, DelBello M, Faist R, Toon T, Clark P, et al. Identifying epilepsy psychiatric comorbidities with machine learning. Acta Neurologica Scandinavica 2020 May;141(5):388-396. [doi: 10.1111/ane.13216] [Medline: 31889296]
83. Guntuku SC. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355. [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
84. Sharma B. Validation of an alcohol misuse classifier in hospitalized patients. Alcohol (Fayetteville, N.Y.) 2020 May;84:49-55. [doi: 10.1016/j.alcohol.2019.09.008] [Medline: 31574300]
85. Barroilhet SA. Characterizing DSM-5 and ICD-11 personality disorder features in psychiatric inpatients at scale using electronic health records. Psychological Medicine 2020 Oct;50(13):2221-2229. [doi: 10.1017/S0033291719002320] [Medline: 31544723]
86. Topaz M, Murga L, Bar-Bachar O, Cato K, Collins S. Alcohol and substance abuse status from clinical notes: The added value of nursing data. Studies in health technologies and informatics 2019 Aug 21;264:1056-1060. [doi: 10.3233/SHTI190386] [Medline: 31438086]
87. Ambalavan AK, Moulahi B, Azé J, Bringay S. Unveiling Online Suicide Behavior: What Can We Learn About Mental Health from Suicide Survivors of Reddit? Studies in health technologies and informatics 2019 Aug;264:50-54. [doi: 10.3233/SHTI190181] [Medline: 31437883]
88. Bittar A, Velupillai S, Roberts A, Dutta R. Text classification to inform suicide risk assessment in electronic health records. Studies in health technologies and informatics 2019 Aug;264:40-44. [doi: 10.3233/SHTI190179] [Medline: 31437881]
89. Yan H, Fitzsimmons-Craft EE. Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention. The International Journal of Eating Disorders 2019 Oct;52(10):1150-1156. [doi: 10.1002/eat.23148] [Medline: 31381168]
90. Danielsen AA. Predicting mechanical restraint of psychiatric inpatients by applying machine learning on electronic health data. Acta Psychiatrica Scandinavica 2019 Aug;140(2):147-157. [doi: 10.1111/acps.13061] [Medline: 31209866]
91. Mezuk B, Ko TM. Suicide Among Older Adults Living in or Transitioning to Residential Long-term Care, 2003 to 2015. JAMA Network Open 2019 Jun;2(6):e195627. [doi: 10.1001/jamanetworkopen.2019.5627] [Medline: 31199445]

XSL•FO
RenderX

92. Parthipan A. Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS ONE 2019 Feb 06;14(2):01-13. [doi: 10.1371/journal.pone.0210575] [Medline: 30726237]
93. Vaucheret Paz E. Sentiment Analysis in Children with Neurodevelopmental Disorders in an Ingroup/Outgroup Setting. Journal of Autism and Developmental Disorders 2020 Jan;50(1):162-170. [doi: 10.1007/s10803-019-04242-3] [Medline: 31571066]
94. Van Le D, Montgomery J, Kirkby KC. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. Journal of Biomedical Informatics 2018 Oct:49-58. [doi: 10.1016/j.jbi.2018.08.007] [Medline: 30118855]
95. Corcoran CM, Carrillo F, Fernandez-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 2018 Feb;17(1):67-75. [doi: 10.1002/wps.20491] [Medline: 29352548]
96. Moessner M, Feldhege J, Wolf M, Bauer S. Analyzing big data in social media: Text and network analyses of an eating disorder forum. International Journal of Eating Disorders 2018 Jul;51(7):656-667. [doi: 10.1002/eat.22878] [Medline: 29746710]
97. Takano K, Ueno M, Mori M, Nishiguchi Y, Raes F. Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. Behavior Research Methods 2017 Jun;49(3):835-852. [doi: 10.3758/s13428-016-0753-x] [Medline: 27338931]
98. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting social anxiety treatment outcome based on therapeutic email conversations. IEEE Journal of Biomedical and Health Informatics 2017 Sep 21;21(5):1449-1459. [doi: 10.1109/JBHI.2016.2601123] [Medline: 27542187]
99. Tung C, Lu W. Analyzing depression tendency of web posts using an event- driven depression tendency warning model. Artificial Intelligence in Medicine 2016 Jan:53-62. [doi: 10.1016/j.artmed.2015.10.003] [Medline: 26616420]
100. Patel R, Jayatilleke N, Broadbent M, Chang CK, Foskett N, Gorrell G, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open 2015 Sep 7;5(9):7619. [doi: 10.1136/bmjopen-2015-007619] [Medline: 26346872]
101. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient recordsinformation in a large mental health case register. PLoS ONE. e013 2015;10(8):4208. [doi: 10.1371/journal.pone.0134208]
102. Patel R, Lloyd T, Jackson R, Ball M, Shetty H, Broadbent M, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. BMJ Open 2015 May 21;5(5):7504. [doi: 10.1136/bmjopen-2014-007504] [Medline: 25998036]
103. Sohn S, Kocher JPA, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. J Am Med Inform Assoc 2011 Dec:i144-9. [doi: 10.1136/amiajnl-2011-000351] [Medline: 21946242]
104. Yu LC, Chan CL, Lin CC, Lin IC. Mining association language patterns using a distributional semantic model for negative life event classification. Journal of Biomedical Informatics 2011 Feb 01;44(4):509-518. [doi: 10.1016/j.jbi.2011.01.006] [Medline: 21292030]
105. Miller WR, Rollnick S. Motivational Interviewing, Helping People Change. New-York: Guilford Publications; 2012:-482.
106. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on twitter: A corpus-based study. J Med Internet Res 2017 Feb;19(2):e48. [doi: 10.2196/jmir.6895] [Medline: 28246066]
107. Insel T, Cuthbert B, Garvey M, Heissen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010 Jul;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
108. Stoyanov D, Telles-Correia D, Cuthbert BN. The Research Domain Criteria (RDoC) and the historical roots of psychopathology: A viewpoint. Eur. psychiatr 2019 Jan 21;57:58-60. [doi: 10.1016/j.eurpsy.2018.11.007] [Medline: 30677549]
109. Ross J, Neylan T, Weiner M, Chao L, Samuelson K, Sim I. Towards constructing a new taxonomy for psychiatry using self-reported symptoms. Stud Health Technol Inform 2015;216:736-740. [Medline: 26262149]
110. Vilar A, Pérez-Sola V, Blasco MJ, Pérez-Gallo E, Ballerster Coma L, Batlle Vila S, et al. Translational research in psychiatry: The research domain criteria project (RDoC). Rev Psiquiatr Salud Ment 2018 Jun;12(3):187-195. [doi: 10.1016/j.rpsm.2018.04.002] [Medline: 29941228]
111. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov;132(20):1920-1930. [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
112. Gholami-Kordkheili F. The impact of social media on medical professionalism: a systematic qualitative review of challenges and opportunities. Journal of Medical Internet Research 2013 Aug;15(8):e184. [doi: 10.2196/jmir.2708] [Medline: 23985172]
113. Fagherazzi G. Digital diabetes: Perspectives for diabetes, prevention and management and research. Diabetes & Metabolism 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
114. Sassolas M. Défense de la psychiatrie clinique. Paris: Eres; 2014.

XSL•FO
RenderX

115. Martinez-Martin N. Is It Ethical to Use Prognostic Estimates from Machine Learning to Treat Psychosis? AMA Journal of Ethics 2018 Sep;20(9):e804-811. [doi: 10.1001/amajethics.2018.804] [Medline: 30242810]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
CEGS N-GRID: Centers of Excellence in Genomic Science Neuropsychiatric Genome-Scale and RDoC Individualized Domains
cTAKES: clinical Text Analysis and Knowledge Extraction System
CUI: concept unique identifier
DSM: Diagnostic and Statistical Manual of Mental Disorders
EHR: electronic health record
EMR: electronic medical record
LDA: latent Dirichlet allocation
LogiR: logistic regression
LSA: latent semantic analysis
ML: machine learning
NLP: natural language processing
NN: neural network
POS: part of speech
PTSD: posttraumatic stress disorder
RDoC: Research Domain Criteria
RF: random forest
SVM: support vector machine
UMLS: unified medical language system

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by M Sharma, H Mufti, M del Pozo Banos, V Curcin, G Hu, D Leightley, SGS Shah, N Miyoshi, F Lanfranchi; comments to author 21.10.19; revised version received 18.04.20; accepted 02.10.20; published 04.05.21.

Please cite as:
Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVylder J, Walter M, Berrouiguet S, Lemey C
Machine Learning and Natural Language Processing in Mental Health: Systematic Review
J Med Internet Res 2021;23(5):e15708
URL: https://www.jmir.org/2021/5/e15708
doi:10.2196/15708
PMID:33944788

©Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Taylor C Ryan, Jonathan Marsh, Jordan DeVylder, Michel Walter, Sofian Berrouiguet, Christophe Lemey. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2021. 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

Review

Impact of Asynchronous Electronic Communication–Based Visits on Clinical Outcomes and Health Care Delivery: Systematic Review

Oliver T Nguyen[1,2], BS; Amir Alishahi Tabriz[3,4], MD, MPH, PhD; Jinhai Huo[1], MD, MSPH, PhD; Karim Hanna[5], MD; Christopher M Shea[6], MA, MPA, PhD; Kea Turner[3,4], MPH, MA, PhD

[1]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States
[2]Department of Health Services Administration, University of Alabama at Birmingham, Birmingham, AL, United States
[3]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States
[4]Department of Oncological Sciences, University of South Florida, Tampa, FL, United States
[5]Department of Family Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, United States
[6]Department of Health Policy and Management, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Kea Turner, MPH, MA, PhD
Department of Health Outcomes and Behavior
Moffitt Cancer Center
12902 USF Magnolia Drive, MRC-CANCONT
Tampa, FL, 33612
United States
Phone: 1 (813) 745 5213
Fax: 1 (813) 745 6525
Email: kea.turner@moffitt.org

Abstract

Background: Electronic visits (e-visits) involve asynchronous communication between clinicians and patients through a secure web-based platform, such as a patient portal, to elicit symptoms and determine a diagnosis and treatment plan. E-visits are now reimbursable through Medicare due to the COVID-19 pandemic. The state of evidence regarding e-visits, such as the impact on clinical outcomes and health care delivery, is unclear.

Objective: To address this gap, we examine how e-visits have impacted clinical outcomes and health care quality, access, utilization, and costs.

Methods: We conducted a systematic review; MEDLINE, Embase, and Web of Science were searched from January 2000 through October 2020 for peer-reviewed studies that assessed e-visits' impacts on clinical and health care delivery outcomes.

Results: Out of 1859 papers, 19 met the inclusion criteria. E-visit usage was associated with improved or comparable clinical outcomes, especially for chronic disease management (eg, diabetes care, blood pressure management). The impact on quality of care varied across conditions. Quality of care was equivalent or better for chronic conditions, but variable quality was observed in infection management (eg, appropriate antibiotic prescribing). Similarly, the impact on health care utilization varied across conditions (eg, lower utilization for dermatology but mixed impact in primary care). Health care costs were lower for e-visits than those for in-person visits for a wide range of conditions (eg, dermatology and acute visits). No studies examined the impact of e-visits on health care access. It is difficult to draw firm conclusions about effectiveness or impact on care delivery from the studies that were included because many used observational designs.

Conclusions: Overall, the evidence suggests e-visits may provide clinical outcomes that are comparable to those provided by in-person care and reduce health care costs for certain health care conditions. At the same time, there is mixed evidence on health care quality, especially regarding infection management (eg, sinusitis, urinary tract infections, conjunctivitis). Further studies are needed to test implementation strategies that might improve delivery (eg, clinical decision support for antibiotic prescribing) and to assess which conditions can be managed via e-visits.

(J Med Internet Res 2021;23(5):e27531) doi:10.2196/27531

XSL•FO
RenderX

KEYWORDS

telemedicine; telehealth; e-visits; electronic visits; digital care; outcome; delivery; review; access; utilization; cost; patient portal; eHealth

Introduction

Telemedicine—or the delivery of health care at a distance—can improve health care access and quality while reducing health care utilization and costs [1]. For instance, telemedicine can improve access to specialists for patients in rural areas or in underresourced care settings by increasing the convenience and availability of health care (eg, extended hours, decreased wait times) [2-4]. Studies have also demonstrated that telemedicine may achieve comparable clinical outcomes to in-person care across a variety of conditions, such as stroke care [5], heart failure [6,7], hepatitis C [8], and diabetes [8]. Telemedicine can also reduce the utilization of in-person care and reduce health care costs [3,4,9]. Furthermore, some telemedicine types have demonstrated cost-effectiveness [10,11]. Studies have shown that the impact of telemedicine on health care delivery and patient outcomes varies across telemedicine types. Some forms of telemedicine, such as telestroke [12-14], have a strong evidence base while other forms of telemedicine, such as electronic visits (e-visits), are understudied.

E-visits involve asynchronous communication between clinicians and patients through a secure web-based platform, such as a patient portal. Generally, patients answer questions about their medical history and symptoms through a structured questionnaire and upload photos (if relevant). The data are then reviewed by a clinician, who develops a diagnosis and treatment plan. Although e-visits often involve clinicians who the patient is familiar with, they can also involve a third-party clinician through direct-to-consumer telemedicine. E-visits offer notable benefits, such as allowing patients and clinicians to communicate at a convenient time (ie, eliminating scheduling barriers) and improving documentation of patient-clinician communication (eg, patients can review clinician's instructions) [15,16]. However, e-visits have a number of implementation barriers, such as those regarding workflow integration (eg, having dedicated clinician time to respond to messages) [16-18], those regarding lack of reimbursement [15,16,19,20], and concerns about the quality of communication from patients (eg, failing to submit sufficient information for diagnosis) [21]. Health care organizations have also raised concerns about the quality of care provided through e-visits, such as the potential for inappropriate antibiotic prescribing and difficulties with providing care without being able to see the patient face-to-face [22-29]. Despite these challenges, studies have reported positive effects of e-visits on health care delivery, such as lower costs [24], comparable follow-up rates to those of in-person care (ie, a proxy for diagnostic accuracy) [30], and comparable or improved patient outcomes (eg, lower uric acid level for patients with gout) [31].

Because of COVID-19, e-visits are being implemented more frequently and are now reimbursable by Medicare and other payers in the United States [32,33]. As the nation moves forward in telemedicine implementation and health care systems decide whether to integrate telemedicine into future care delivery, it is critical to determine the costs and benefits of telemedicine models, such as e-visits. To date, a systematic review has not been conducted to assess the state of the evidence regarding e-visits. To address this gap, the objective of this review is to summarize the findings of studies examining the association of e-visits with clinical outcomes, quality of care, access to care, utilization, and costs.

Methods

We conducted a systematic review based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) guidelines [34].

Data Sources and Searches

MEDLINE, Embase, and Web of Science were searched to locate peer-reviewed studies published from January 2000 through October 2020. The start date for the search was chosen because electronic visits were developed recently, and the authors did not anticipate any studies published prior to 2000 (when web-based, patient portal studies emerged). A preliminary search was done to confirm this. Studies were limited to those in the United States as systemic factors, such as reimbursement, may influence the results. All reference lists for included studies were cross-searched. Duplicate studies were removed. Multimedia Appendix 1 lists the search terms used for this search after consultation with a health sciences librarian at the University of Florida.

Study Selection

Two reviewers (KT, OTN) independently screened papers identified from the search strategy, and assessed each for inclusion eligibility using a spreadsheet (Excel 2013, Microsoft Inc). Any discrepancy encountered was discussed until a consensus was reached.

E-visits were defined as any asynchronous electronic visit where a clinician assesses a patient's health status, makes a diagnosis, and develops a treatment plan via a secure messaging system (eg, patient portal) [24,35]. This definition was based on how recent studies have defined e-visits [24,35]. Studies that defined an e-visit differently were excluded (eg, defines e-visit as real-time, 2-way communication). Studies were included that reported on the impact of e-visits on clinical outcomes, health care quality, access, utilization or costs. Included studies also had to be written in English, empirical (ie, reporting original research), quantitative, and peer reviewed.

Data Extraction and Quality Assessment

For each included study, 2 investigators (KT, OTN) noted study design, outcome measurements, care setting, medical conditions studied, sample size, and major findings. Since most of the studies included were observational, the authors used the Risk of Bias Assessment Tool for Nonrandomized Studies to assess criteria specific to observational studies [36]. A *P*-value <.05 was considered significant.

XSL•FO
RenderX

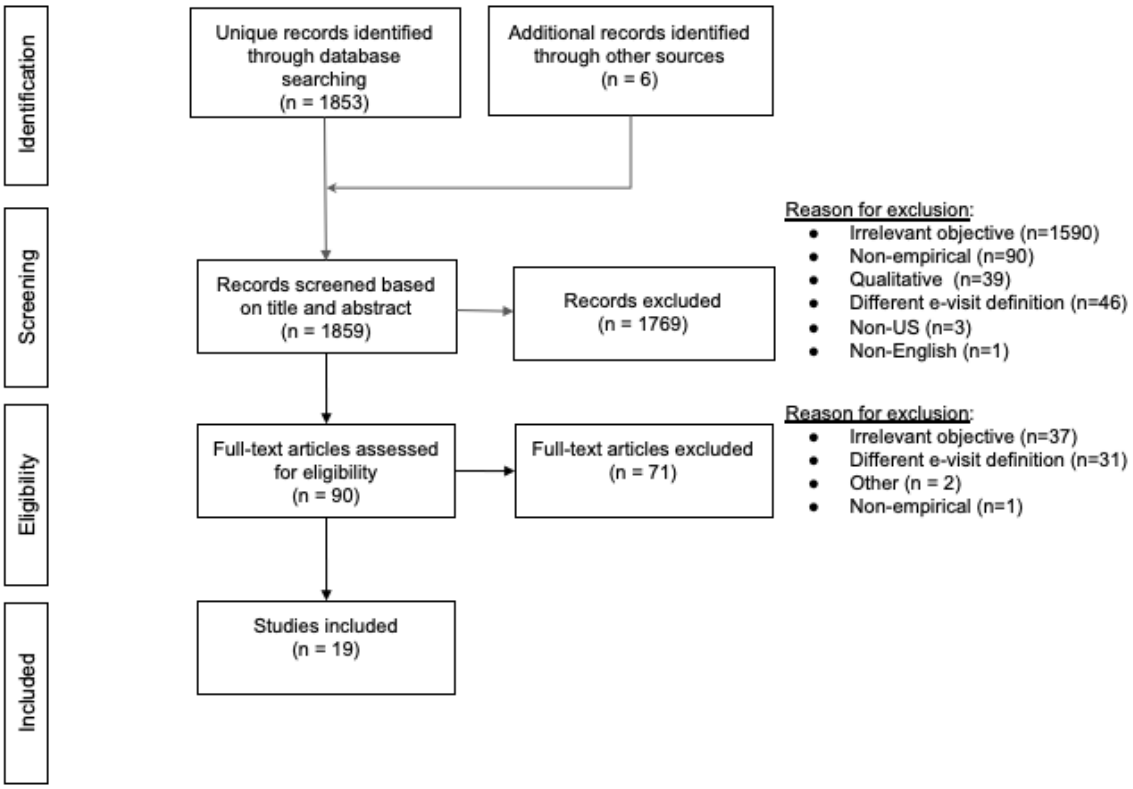
Data Synthesis and Analysis

Due to the heterogeneity of outcome measures used in our included studies, it was infeasible to conduct a meta-analysis. Consequently, findings were qualitatively grouped by outcome type (eg, clinical outcomes, costs).

Results

After reviewing 1859 studies, a total of 19 studies met our inclusion criteria. Figure 1 shows our study selection process.

Figure 1. PRISMA [34] flowchart.



Study Characteristics

Most studies assessed clinical outcomes, quality of care, health care utilization, and costs. None reported access outcomes. Most studies were observational in design, with one study using a randomized experimental design, another study using a case study design, and another 2 studies employing quasi-experimental designs. Sixteen studies used cross-sectional or pooled cross-sectional data. Three studies used a panel design. Table 1 summarizes the study characteristics of included studies, and Table 2 summarizes findings with respect to clinical outcomes, quality of care, access to care, health care utilization, and costs.

XSL•FO
RenderX

Table 1. Study characteristics.

Citation	Types of health system measure	Entity provisioning e-visits	Care setting	Medical conditions treated	Sample size, n
Adamson et al [19]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	Sinusitis, depression, back pain, cough, anxiety, hypertension, abdominal pain, headache, urinary tract infections, influenza, allergic rhinitis, dermatitis, attention-deficit/hyperactive disorder, gastroesophageal reflux disease, vaginitis, upper respiratory infection, insomnia, asthma, contraception, hyperlipidemia	2531
Rohrer et al [37]	Costs, utilization	Health care organization	Outpatient clinic	Conjunctivitis, sore throat, viral illness, bronchitis, cough	390
Watson et al [38]	Clinical outcomes	Health care organization	Outpatient clinic	Acne	121
Albert et al [39]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	Blood pressure management, fractures, diabetes management, skin conditions, prostatitis, pain, sleep issues, vomiting, mononucleosis, hemorrhoids, cold symptoms	121
Courneya et al [40]	Quality of care, costs, utilization	Insurer	N/A[a]	Acute sinusitis, chronic sinusitis, urinary tract infections, conjunctivitis, viral upper respiratory infection, lower genitourinary system infection, yeast infection, otolaryngology diseases, acne, allergic rhinitis, acute bronchitis	Reported as more than 40,000
Mehrotra et al [24]	Quality of care, costs, utilization	Health care organization	Outpatient clinic, academic medical center	Sinusitis, urinary tract infections	574
North et al [41]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	No specific conditions studied, but e-visit content mentioning chest pain and respiratory complications were monitored	892
Heyworth et al [42]	Quality of care	Health care organization	Inpatient, Veterans Affairs	Diabetes, hypertension, hyperlipidemia, heart disease, prior history of myocardial infarction or stroke	51
Pathipati et al [43]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	Rash, acne, other unspecified dermatological conditions	38
Hawes et al [44]	Clinical outcomes, quality of care	Health care organization	Outpatient clinic, academic medical center	Diabetes, anticoagulation management	36
Levine et al [45]	Clinical outcomes, costs, utilization	Health care organization	Outpatient clinic, academic medical center	Hypertension	1786
Penza et al [46]	Quality of care, costs, utilization	Health care organization	Outpatient clinic, academic medical center	Conjunctivitis	505
Penza et al [47]	Clinical outcomes, costs, utilization	Health care organization	Outpatient clinic, academic medical center	Allergies, upper respiratory infection, cold sores, influenza, lice, conjunctivitis, sinusitis, sore throat, sunburn, tick exposure, urinary tract infections, yeast infection	1009
Player et al [48]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	Athlete's foot, allergic skin reaction, eczema, cold sore, shingles rash, marine animal string, jock itch, nosebleed, poison ivy, rash, red eye, ringworm, scabies, hemorrhoids, sunburn, back pain, gout, heartburn, seasonal allergies, travel precaution, prescription refills, sexually transmitted infections, diarrhea, influenza, sinus problems, urinary problems, vaginal irritation or discharge	1565
Rajda et al [49]	Costs, utilization	Direct-to-consumer company	N/A	Acne vulgaris, atopic dermatitis, onychomycosis, psoriasis vulgaris, rosacea	395

XSL•FO
RenderX

Citation	Types of health system measure	Entity provisioning e-visits	Care setting	Medical conditions treated	Sample size, n
Hertzog et al [30]	Costs, utilization	Health care organization	Outpatient clinic, academic medical center	Yeast infection, acne, allergic rhinitis, urinary tract infections, upper respiratory infection, conjunctivitis, oral sores, irritable bowel syndrome, tobacco cessation	2691
Murray et al [50]	Clinical outcomes, quality of care	Health care organization	Outpatient clinic, academic medical center	Urinary tract infections	300
Penza et al [51]	Clinical outcomes, quality of care	Health care organization	Outpatient clinic, academic medical center	Acute sinusitis	300
Yokose et al [31]	Quality of care	Health care organization	Outpatient clinic, academic medical center	Gout	124

[a]N/A: not applicable.

XSL•FO
RenderX

Table 2. Effects of asynchronous e-visits on clinical outcomes, quality, utilization, and costs.

Outcome type and citations	Specific measures	Impact
Clinical outcomes		
Hawes et al [44]	Abnormal international normalized ratio	E-visits were associated with lower abnormal international normalized ratios than in-person visits (P<.05).
Hawes et al [44]	HbA_{1c}[a]	E-visits were associated with lower HbA_{1c} values than in-person visits (P<.001).
Hawes et al [44]	Amount of diabetic patients with controlled blood pressure	Compared to the preimplementation period, more diabetic patients were observed with controlled blood pressure in the postimplementation period (P<.001).
Hawes et al [44]	Amount of diabetic patients with HbA_{1c} levels of less than 8%	Compared to the preimplementation period, more diabetic patients with HbA_{1c} levels of less than 8% were observed in the postimplementation period (P<.0001).
Hawes et al [44]	Amount of diabetic patients with HbA_{1c} levels of less than 7%	Compared to the preimplementation period, more diabetic patients with HbA_{1c} levels of less than 7% were observed in the postimplementation period (P<.001).
Levine et al [45]	Systolic blood pressure	Equivalent outcomes
Watson et al [38]	Total inflammatory lesion counts	Equivalent outcomes
Watson et al [38]	Frontal inflammatory lesion counts	Equivalent outcomes
Watson et al [38]	Leeds score	Equivalent outcomes
Penza et al [47,51]	Mortality rate	Penza et al reported only descriptive statistics in both studies, so it is unclear if there are differences in mortality rates between e-visits and in-person visits.
Murray et al [50]; Penza et al [51]	Hospitalizations	Murray et al [50] and Penza et al [51] reported only descriptive statistics, so it is unclear if there are differences in the number of related hospitalizations between e-visits and in-person visits.
Murray et al [50]	Antibiotic retreatment rate	Equivalent outcomes
Yokose et al [31]	Proportion of patients serum urate levels of less than 6.0 mg/dL	E-visits had greater proportions of patients with optimal control of serum urate levels when compared to in-person visits (P<.01).
Quality of care		
Hawes et al [44]	Amount of diabetic patients receiving aspirin, if clinically indicated	Equivalent outcomes
Hawes et al [44]	Amount of diabetic patients receiving moderate-intensity statins	Equivalent outcomes
Hawes et al [44]	Amount of diabetic patients receiving high-intensity statins	Equivalent outcomes
Heyworth et al [42]	Medication discrepancy discovery rate	It is unclear what the impact is on the rate of discovering medication discrepancies as no P value was reported.
Mehrotra et al [24]	Order rate of diagnostic test	E-visits had a lower order rate of diagnostic tests when compared to in-person visits (P<.001).
Mehrotra et al [24]	Order rate of preventive care services	E-visits had a lower order rate of preventive care services when compared to in-person visits (P<.01).
Yokose et al [31]	Rate that serum urate levels were checked	E-visits had more frequent checks of serum urate levels when compared to in-person visits (P<.05).
Murray et al [50]; Penza et al [46,51]; Mehrotra et al [24]; Courneya et al [40]	Antibiotic prescribing rate	Mehrotra et al [24] reported that the rate of prescribing antibiotics was higher during e-visits than in-person visits for sinusitis (P<.001) but not for e-visits for urinary tract infections. However, Penza et al [46] and Murray et al [50] saw equivalent outcomes. Penza et al [51] reported e-visits had lower antibiotic prescribing rates than in-person visits (P<.001). Courneya et al [40] also investigated the association but did not report a P value, so the impact on antibiotic prescribing rate is unclear.
Health care utilization		
Levine et al [45]	Overall primary care visit utilization	Equivalent outcomes

XSL•FO
RenderX

Outcome type and citations	Specific measures	Impact
Levine et al [45]	Overall specialist visit utilization	Equivalent outcomes
Levine et al [45]	Overall emergency department utilization	Equivalent outcomes
Levine et al [45]	Overall inpatient admissions	Equivalent outcomes
Rajda et al [49]	Number of specialist procedures done 60- and 90-days after initial consultation	E-visits were associated with a lower number of specialist procedures performed 60 and 90 days after an initial consultation when compared to in-person visits (P<.01)
Murray et al [50]; Penza et al [51]	30-day follow-up rate (planned and unplanned)	Equivalent outcomes
Penza et al [46]; Pathipati et al [43]; Albert et al [39]; Player et al [48]; Adamson et al [19]	Rate of patients who need planned follow-up visits	Penza et al [46] reported that e-visits were associated with higher rates of planned follow-up visits than in-person visits (P<.001). Pathipathi et al [43], Albert et al [39], Player et al [48], and Adamson et al [19] reported only the proportion of e-visits that required follow-up visits, so it is unclear what the association of e-visit usage and rate of planned follow-up visits is.
Penza et al [47]; Mehrotra et al [24]; Courneya et al [40]; North et al [41]; Hertzog et al [30]	Unexpected follow-up encounter rate after initial encounter	Hertzog et al [30] reported that e-visits were associated with higher unexpected follow-up rates when compared to in-person visits (P<.05). However, Mehrotra et al [24], Courneya et al [40], and North et al [41] found equivalent outcomes. Penza et al [47] reported only descriptive statistics, so it is unclear if there are differences in unexpected follow-up encounter rates between e-visits and in-person visits.
Health care costs		
Rajda et al [49]; Courneya et al [40]; Rohrer et al [37]	Treatment costs	Courneya et al [40] and Rajda et al [49] reported e-visits were associated with lower treatment costs (P<.001). Rohrer et al [37] reported a lower median of costs associated with e-visits than in-person visits (P<.01).

[a]HbA_{1c}: hemoglobin A_{1c}.

Study Quality Assessment Results

Detailed results of the quality assessment are summarized in Multimedia Appendix 2. Each column describes the quality criterion we assessed. Briefly, most (12/19, 63.2%) studies reported a strategy for minimizing selection bias. Less than half of studies (9/19, 47.4%) reported methods to control for confounders. All studies measured outcomes for e-visits separately from other forms of telemedicine, making it possible to evaluate the unique impact of e-visits on the outcome of interest. Problems with low response rate (<50%) or attrition bias (>10% dropout) were less common (2/19, 10.5%).

Clinical Outcomes

Associations between e-visit usage and clinical outcomes were reported by 7 studies. A total of 13 different measures were used among the studies. Overall, the studies reported an association with improved outcomes or null findings across the medical conditions examined.

Among diabetic patients, one study (n=36 patients) found that e-visits were associated with significantly improved glucose over a 6-month period (–3.4 percentage points in HbA_{1c}, P<.001) [44]. In the same study, e-visits for anticoagulant management were associated with less frequent instances of abnormal international normalized ratio values compared to in-person care (5/104, 5% vs 1/198, 0.5%, P<.05) [44]. An additional study (n=62) reported that patients with gout who received care through e-visits were more likely to have optimal serum rate levels (>6.0 mg/dL) (63.8% vs 33.9%, P<.01) and lower mean serum urate levels (5.5 mg/dL vs 6.7 mg/dL, P<.01) compared to historical controls [31]. In the context of acne and hypertension management, equivalent outcomes were reported between e-visits and in-person visits [38,45].

Quality of Care

Across the 8 studies that examined the association of e-visit usage on quality of care, 8 unique measures were observed. Overall, there were mixed effects on the association of e-visit usage on quality of care.

Mixed results were observed in some quality of care measures of several health care conditions between e-visits and in-person visits [44]. For example, one study (n=36) [44] reported comparable prescribing rates for statins among patients with diabetes across e-visits and in-person visits. One study [24] found that e-visit usage was associated with significantly lower rates of diagnostic procedures for sinusitis (0/475, 0% vs 40/4690, 1%, P=.04) and lower rates of diagnostic procedures for urinary tract infections (8/99, 8% vs 1501/2855, 53%, P<.001) than those for historical controls. The same study [24] also reported that preventive screenings were lower among sinusitis patients receiving an e-visit (1/475, 0.2% vs 155/4690, 3%; P<.001) and urinary tract infection patients' receiving an e-visit (0/99, 0% vs 214/2855, 7%, P=.005) than in those receiving in-person visits.

When examining management of acute infections (eg, urinary tract infections), the definitive impact was less clear. One study [24] found that e-visits resulted in higher antibiotic prescribing rates for sinusitis (471/475, 99% vs 4408/4690, 94%, P<.001)

XSL•FO
RenderX

but no association for urinary tract infections (98/99, 99% vs 1299/2855, 92%, *P*=.07), compared to those for historical control. Another study [46] reported that e-visits for conjunctivitis resulted in a significantly lower antibiotic prescribing rate than that of phone visits (26/101, 26% vs 84/202, 42%, *P*=.006), and a third study [51] found that e-visits for sinusitis resulted in a significantly lower antibiotic prescribing rate (84/150, 56% vs 108/150, 72%, *P*=.01) than that of in-person visits. Lastly, one study (n=450) [50] reported no difference in antibiotic prescribing rates for treatment of urinary tract infections across e-visits, phone encounters, and in-person visits.

Access to Care

No studies investigated the association between e-visits and access to care.

Health Care Utilization

Fourteen studies assessed health care utilization associated with e-visits. An overall mixed impact was observed across these studies.

Some studies evaluated the impact of e-visits on subsequent health care utilization, such as primary care, specialty care, and emergency care [45,49-51]. For example, one study [45] matched 893 e-visits and 893 in-person visits for hypertension and found that e-visits resulted in fewer primary care visits (–0.8 visits, 95% CI 0.3-1.2) compared to in-person visits. The same study [45] also found that usage of specialist visits, emergency department visits, and inpatient admissions were not significantly different across in-person and e-visits. Another study [49] evaluating a teledermatology program found that e-visits were associated with significantly fewer specialty visits at 60-day (15 vs 46, *P*=.005) and 90-day follow-up (26 vs 74, *P*=.001) compared to those associated with in-person visits. Another study [40] examining a direct-to-consumer telemedicine program found that the rate of visits that did not require a follow-up visit (ie, resolution rate) was similar for e-visits for sinusitis (90% vs 91%) and conjunctivitis (94% vs 95%) to those for in-person visits.

Some studies [24,40,41] reported no differences in health care utilization. For example, one study [24] found that e-visits for sinusitis and urinary tract infections had equivalent rates of follow-up visits, phone calls, and emails within 3-week follow-up. Some studies [30,46] reported a higher rate of health care utilization with e-visits. For example, one study [30] found that the rate of follow-up visits was higher for primary care e-visits (59/490, 12%) compared to that for in-person visits (198/2201, 9%; *P*=.04).

Health Care Costs

Three studies [37,40,49] found e-visit usage was associated with lower overall treatment costs than those for in-person visits. One study utilizing claims data found that e-visits for dermatology resulted in lower mean costs at the initial visit (US $59 vs $113, *P*<.001), at 30-day follow-up ($70 vs $202, *P*=.03), and at 60-day follow-up ($78 vs $221, *P*=.02) than those of in-person visits; however, at the 90-day follow-up, costs were comparable to those of an in-person visit ($86 vs $307, *P*=.08) [49]. One study [40] of a direct-to-consumer e-visit platform reported lower costs for a wide-range of conditions, including sinusitis, conjunctivitis, acne, and ear, nose, and throat infections. As an example, that study (n=9551 visits) used claims data to report that the cost per visit for e-visits was significantly lower for acne management ($178 vs $361, *P*<.001) compared to in-person care [40]. A third study [37] calculated total reimbursable costs and descriptively reported the median cost per visit was lower for e-visits ($161 vs $219) compared to in-person visits for acute conditions (eg, conjunctivitis, sore throat, bronchitis, viral illness, cough).

Discussion

This was a systematic review that assessed the impact of e-visits on clinical outcomes and health care delivery. To our knowledge, this is the first systematic review to assess the state of evidence for asynchronous e-visits. Most studies found that e-visits were associated with lower treatment costs and comparable clinical outcomes to in-person visits. Studies reported mixed effects on health care utilization, and no studies evaluated the impact of e-visits on health care access, suggesting future research is needed in this area. We provide implications for research and practice below.

Our review found that e-visits may be an adequate substitute for in-person care for chronic disease management. Studies showed that e-visits were effective for the treatment of diabetes, hypertension, and gout. Our findings are consistent with those of previous studies [52] that suggest that other forms of asynchronous communication with providers (eg, secure messaging without a formal e-visit) can improve chronic disease management. Further study, however, is needed to determine the ideal conditions for e-visit implementation. For example, many chronic conditions co-occur with other conditions (eg, diabetes and hypertension), and current studies were not designed to determine whether e-visits are effective for complex patients, such as patients with multiple chronic conditions and older adults who are frail. Additionally, it is important to note that not all types of medical issues may be appropriate for asynchronous management and health care systems may need to implement safeguards to ensure that the right patient is using an e-visit [53]. For example, if patients send a message about an urgent condition, e-visit technology could be harnessed to flag certain keywords (eg, chest pain, breathing difficulties) and display automated pop-up alerts to patients to instruct them to seek care in-person at the office or emergency department [54]. Some health care systems have also implemented guidelines, such as requiring at least one in-person visit prior to an e-visit to ensure that a patient has an established relationship with a clinician [55]. Additional research is needed to determine how e-visits should be implemented (eg, how to prevent inappropriate usage and for which patients does it work best).

Our review found mixed evidence for the effect of e-visits on quality of care. For example, prior studies [22,24-29] have raised concerns that telemedicine usage can increase inappropriate antibiotic prescribing, but this review found mixed evidence on whether comparable antibiotic prescribing rates were observed between e-visits and in-person visits. This inconsistency may

XSL•FO
RenderX

stem from differences in the acute infection that was studied (ie, acute sinusitis, urinary tract infections, ear infections, conjunctivitis). Further research is needed to better understand when e-visits can be used effectively for managing acute infections and what implementation strategies can be used to ensure appropriate antibiotic prescribing (eg, use of clinical decision support). Furthermore, additional research is needed to compare the receipt of low-value care (eg, overutilization of services for sinusitis and urinary tract infections across e-visits and in-person visits). Future studies should examine guidelines, such as the Choosing Wisely guidelines [56,57], to see whether e-visits reduce low-value care compared to in-person visits. One study [24] included in this review found that e-visits were associated with lower rates of diagnostic procedures for sinusitis and urinary tract infections and may offer an advantage in terms of reducing unnecessary health care utilization. Studies in this review did report other quality problems, such as lower use of preventive care services in e-visits compared to in-person visits. This may stem from a lack of practice guidance on how e-visits should be implemented (eg, should providers use e-visits as an opportunity to reinforce messages about preventive care?). Prior studies [16] have noted that lack of practice guidance or standard procedures as barriers to e-visit implementation. These differences in quality of care suggest additional implementation research is needed to test implementation strategies for ensuring quality of care delivered through e-visits is consistent across clinicians.

Our review found that e-visits had mixed effects on health care utilization compared to in-person visits. These findings are consistent with those of reviews on other forms of telemedicine [6,58-62]. Specifically, telemedicine can reduce health care utilization in certain instances (eg, reduce the need for in-person visits) or it can increase utilization (eg, meet an unmet demand for a patient that was not previously accessing in-person care). Variation in health care utilization may stem from other various factors that are unaccounted for. For example, if a patient did not submit sufficient information to be evaluated through an e-visit, this could lead to greater health care utilization compared to an initial in-person visit where information exchange is synchronous. Future research is needed to determine strategies for ensuring that complete information is elicited from the patient (eg, structured symptom questionnaires). Additional research should test strategies for optimizing patient data collection and patient-clinician communication through e-visit platforms.

The findings on favorable cost implications align with those of other works that evaluated the effects that synchronous alternatives for care (eg, telemedicine via teleconferencing software) had on cost outcomes for patients [63-70]. Notably, 2 [40,49] of 3 studies used claims data to estimate health care costs; however, insurers commonly include in their contracts with health care organizations contractual adjustments (ie, slightly lower reimbursement rates in exchange for including the health care organization in-network) [71], suggesting that using claims data may underestimate the true cost savings potential of e-visits. Furthermore, many of these studies did not evaluate cost comprehensively (eg, only examined costs of dermatology visits and not visits to providers outside of that delivery system, analyzing average costs per visit instead of average costs of an episode of illness). Additionally, we did not identify any studies that assessed cost-effectiveness, which considers costs relative to outcomes. Future cost-effectiveness studies may also help health care systems make decisions about whether e-visits are worth the investment.

Additional research is also needed on how e-visits impact health care access and the digital divide. There is growing evidence that patient-level disparities exist across adoption and usage patterns of patient portals [72-76]. Similarly, several studies in this review reported differences in usage based on sex [19,30,39,41,42,48], age [30,42,47], employment status [41,48], and ethnicity [19,41]. Since many patients access e-visits through the patient portal, which has known disparities in uptake [72-76], future studies are needed to test strategies for overcoming disparities in patient portal adoption. Since the Pew Research Center reports 81% of Americans in 2019 owned a smartphone [77], there have been studies that have recently tested whether smartphone access to a patient portal could improve access [78,79]. Future studies should test whether strategies, such as smartphone access, could improve uptake of e-visits.

Some of the included literature cited implementation barriers in their discussion that should be further explored in future studies (eg, lack of integration into workflow). Clinicians may need assistance with adjusting their workflows when being trained on e-visits [80]. Best practices should be researched and disseminated to help alleviate concerns on increased workload [81]. Furthermore, the adoption of payment models by insurers that reimburse for e-visits may be a crucial facilitator of e-visit uptake [20,82,83]. Since the COVID-19 pandemic has spurred the introduction of insurance coverage of e-visits among Medicare and some private payers, case studies have been published that suggest e-visits are being used more frequently during the COVID-19 pandemic. For example, one health care system reported that the use of e-visits increased by 4000% during the COVID-19 pandemic and that the majority of e-visits were used for remotely managing patients with COVID-19 [84-86]. Further work is needed to evaluate if the reimbursement policies have led to higher utilization of e-visits by patients and health care organizations.

This systematic review has several limitations. First, a majority of the studies used an observational design, limiting our ability to draw causal conclusions on the effect of using e-visits on quality of care, access to care, costs or clinical outcomes. Second, we found heterogeneity in how studies measured the impact of e-visits, making it impossible to quantitatively pool study estimates. Third, we excluded non-English studies, which may limit our ability to determine the effect of e-visits outside of English-speaking regions. Fourth, the follow-up periods among the studies varied (eg, 2 weeks vs 1 month), limiting our comparisons of findings between studies. Fifth, a majority of the studies examined e-visits that are hosted by the health care delivery organization rather than by a direct-to-consumer vendor or insurer, limiting the generalizability of our results to e-visit programs sponsored by entities external to the health care delivery organization. Lastly, most e-visits occurred in outpatient

XSL•FO
RenderX

ambulatory and academic medical center contexts, limiting our ability to comment on inpatient settings.

Overall, the evidence suggests that e-visits can provide equivalent outcomes to in-person care and reduce health care costs for certain health care conditions. There are still notable quality concerns (eg, inappropriate antibiotic prescribing, underutilization of preventive care) that warrant further study. It is also unknown how e-visits have affected access to care. Furthermore, many studies in the review lacked a rigorous design, making it difficult to draw firm conclusions about the state of evidence regarding e-visits. Future trials should be conducted to test the effectiveness of e-visits and determine what factors drive effective implementation.

Acknowledgments

We owe many thanks to Nancy Schaefer for providing her expertise as a health science librarian by providing us a framework on how to search for relevant articles and suggesting search terms to supplement our original search strategy.

Authors' Contributions

KT conceptualized the study. OTN and KT drafted the manuscript. All authors participated in the analysis and interpretation of data. AAT, JH, KH, CMS, and KT provided critical revisions to the manuscript. All authors approved the submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 8 KB - jmir_v23i5e27531_app1.docx]

Multimedia Appendix 2
Quality assessment of studies.
[DOCX File , 8 KB - jmir_v23i5e27531_app2.docx]

References

1. The Role of Telehealth in an Evolving Health Care Environment: Workshop Summary. Washington, DC: The National Academies Press; Nov 20, 2012.
2. Zhang D, Wang G, Zhu W, Thapa JR, Switzer JA, Hess DC, et al. Expansion of telestroke services improves quality of care provided in super rural areas. Health Aff (Millwood) 2018 Dec;37(12):2005-2013. [doi: 10.1377/hlthaff.2018.05089] [Medline: 30633675]
3. Albritton J, Maddox L, Dalto J, Ridout E, Minton S. The effect of a newborn telehealth program on transfers avoided: a multiple-baseline study. Health Aff (Millwood) 2018 Dec;37(12):1990-1996. [doi: 10.1377/hlthaff.2018.05133] [Medline: 30633672]
4. Anderson D, Villagra VG, Coman E, Ahmed T, Porto A, Jepeal N, et al. Reduced cost of specialty care using electronic consultations for Medicaid patients. Health Aff (Millwood) 2018 Dec;37(12):2031-2036. [doi: 10.1377/hlthaff.2018.05124] [Medline: 30633678]
5. Johansson T, Wild C. Telemedicine in acute stroke management: systematic review. Int J Technol Assess Health Care 2010 Apr;26(2):149-155. [doi: 10.1017/S0266462310000139] [Medline: 20392317]
6. Pekmezaris R, Tortez L, Williams M, Patel V, Makaryus A, Zeltser R, et al. Home telemonitoring in heart failure: a systematic review and meta-analysis. Health Aff (Millwood) 2018 Dec;37(12):1983-1989. [doi: 10.1377/hlthaff.2018.05087] [Medline: 30633680]
7. Lin M, Yuan W, Huang T, Zhang H, Mai J, Wang J. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Dec;65(5):899-911. [doi: 10.1136/jim-2016-000199] [Medline: 28330835]
8. Zhou C, Crawford A, Serhal E, Kurdyak P, Sockalingam S. The impact of project ECHO on participant and patient outcomes: a systematic review. Acad Med 2016 Oct;91(10):1439-1461. [doi: 10.1097/ACM.0000000000001328] [Medline: 27489018]
9. Shah SJ, Schwamm LH, Cohen AB, Simoni MR, Estrada J, Matiello M, et al. Virtual visits partially replaced in-person visits in an ACO-based medical specialty practice. Health Aff (Millwood) 2018 Dec;37(12):2045-2051. [doi: 10.1377/hlthaff.2018.05105] [Medline: 30633681]
10. Switzer JA, Demaerschalk BM, Xie J, Fan L, Villa KF, Wu EQ. Cost-effectiveness of hub-and-spoke telestroke networks for the management of acute ischemic stroke from the hospitals' perspectives. Circ Cardiovasc Qual Outcomes 2013 Jan 01;6(1):18-26. [doi: 10.1161/CIRCOUTCOMES.112.967125] [Medline: 23212458]

XSL•FO
RenderX

11. Nelson RE, Saltzman GM, Skalabrin EJ, Demaerschalk BM, Majersik JJ. The cost-effectiveness of telestroke in the treatment of acute ischemic stroke. Neurology 2011 Oct 25;77(17):1590-1598 [FREE Full text] [doi: 10.1212/WNL.0b013e318234332d] [Medline: 21917781]
12. Baratloo A, Rahimpour L, Abushouk AI, Safari S, Lee CW, Abdalvand A. Effects of telestroke on thrombolysis times and outcomes: a meta-analysis. Prehosp Emerg Care 2018 Jan 18:1-13. [doi: 10.1080/10903127.2017.1408728] [Medline: 29345529]
13. Rubin MN, Wellik KE, Channer DD, Demaerschalk BM. Systematic review of telestroke for post-stroke care and rehabilitation. Curr Atheroscler Rep 2013 Aug;15(8):343. [doi: 10.1007/s11883-013-0343-7] [Medline: 23761014]
14. Kepplinger J, Barlinn K, Deckert S, Scheibe M, Bodechtel U, Schmitt J. Safety and efficacy of thrombolysis in telestroke: a systematic review and meta-analysis. Neurology 2016 Sep 27;87(13):1344-1351. [doi: 10.1212/WNL.0000000000003148] [Medline: 27566746]
15. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]
16. Hickson R, Talbert J, Thornbury WC, Perin NR, Goodin AJ. Online medical care: the current state of "eVisits" in acute primary care delivery. Telemed J E Health 2015 Feb;21(2):90-96. [doi: 10.1089/tmj.2014.0022] [Medline: 25474083]
17. Lieu TA, Altschuler A, Weiner JZ, East JA, Moeller MF, Prausnitz S, et al. Primary care physicians' experiences with and strategies for managing electronic messages. JAMA Netw Open 2019 Dec 02;2(12):e1918287 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18287] [Medline: 31880798]
18. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
19. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710 [FREE Full text] [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
20. Tang PC, Black W, Young CY. Proposed criteria for reimbursing eVisits: content analysis of secure patient messages in a personal health record system. AMIA Annu Symp Proc 2006:764-768 [FREE Full text] [Medline: 17238444]
21. Yau G, Williams A, Brown J. Family physicians' perspectives on personal health records: qualitative study. Can Fam Physician 2011 May;57(5):e178-e184 [FREE Full text] [Medline: 21642732]
22. Ray KN, Shi Z, Gidengil CA, Poon SJ, Uscher-Pines L, Mehrotra A. Antibiotic prescribing during pediatric direct-to-consumer telemedicine visits. Pediatrics 2019 May;143(5):e20182491 [FREE Full text] [doi: 10.1542/peds.2018-2491] [Medline: 30962253]
23. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287. [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
24. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. JAMA Intern Med 2013 Jan 14;173(1):72-74 [FREE Full text] [doi: 10.1001/2013.jamainternmed.305] [Medline: 23403816]
25. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Antibiotic prescribing for acute respiratory infections in direct-to-consumer telemedicine visits. JAMA Intern Med 2015 Jul;175(7):1234-1235. [doi: 10.1001/jamainternmed.2015.2024] [Medline: 26011763]
26. Smith KL, Tran D, Westra BL. Sinusitis treatment guideline adherence in the e-visit setting: a performance improvement project. Appl Clin Inform 2016;7(2):299-307 [FREE Full text] [doi: 10.4338/ACI-2015-10-CR-0143] [Medline: 27437042]
27. Foster CB, Martinez KA, Sabella C, Weaver GP, Rothberg MB. Patient satisfaction and antibiotic prescribing for respiratory infections by telemedicine. Pediatrics 2019 Sep;144(3):e20190844. [doi: 10.1542/peds.2019-0844] [Medline: 31371464]
28. Martinez KA, Rood M, Jhangiani N, Kou L, Boissy A, Rothberg MB. Association between antibiotic prescribing for respiratory tract infections and patient satisfaction in direct-to-consumer telemedicine. JAMA Intern Med 2018 Nov 01;178(11):1558-1560 [FREE Full text] [doi: 10.1001/jamainternmed.2018.4318] [Medline: 30285050]
29. Sprecher E, Finkelstein JA. Telemedicine and antibiotic use: one click forward or two steps back? Pediatrics 2019 Sep;144(3):e20191585. [doi: 10.1542/peds.2019-1585] [Medline: 31371465]
30. Hertzog R, Johnson J, Smith J, McStay FW, da Graca B, Haneke T, et al. Diagnostic accuracy in primary care e-visits: evaluation of a large integrated health care delivery system's experience. Mayo Clin Proc 2019 Jun;94(6):976-984. [doi: 10.1016/j.mayocp.2019.02.011] [Medline: 31171135]
31. Yokose C, Jorge A, D'Silva K, Serling-Boyd N, Matza M, Nasrallah M, et al. Using electronic visits (e-visits) to achieve goal serum urate levels in patients with gout in a rheumatology practice: a pilot study. Semin Arthritis Rheum 2020 Dec;50(6):1382-1386. [doi: 10.1016/j.semarthrit.2020.03.001] [Medline: 32359694]
32. Medicare telemedicine health care provider fact sheet. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2021-04-24]
33. Use of communication technology-based services during coronavirus/COVID-19 Internet. American Speech-Language-Hearing Association. 2021 Jan 29. URL: https://www.asha.org/Practice/reimbursement/medicare/Use-of-E-Visit-Codes-for-Medicare-Part-B-Services-During-Coronavirus/ [accessed 2021-04-24]

XSL•FO
RenderX

34. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
35. Mehrotra A. The convenience revolution for treatment of low-acuity conditions. JAMA 2013 Jul 3;310(1):35-36 [FREE Full text] [doi: 10.1001/jama.2013.6825] [Medline: 23821082]
36. Kim SY, Park JE, Lee YJ, Seo H, Sheen S, Hahn S, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. J Clin Epidemiol 2013 Apr;66(4):408-414. [doi: 10.1016/j.jclinepi.2012.09.016] [Medline: 23337781]
37. Rohrer JE, Angstman KB, Adamson SC, Bernard ME, Bachman JW, Morgan ME. Impact of online primary care visits on standard costs: a pilot study. Popul Health Manag 2010 Apr;13(2):59-63. [doi: 10.1089/pop.2009.0018] [Medline: 20415617]
38. Watson AJ, Bergman H, Williams CM, Kvedar JC. A randomized trial to evaluate the efficacy of online follow-up visits in the management of acne. Arch Dermatol 2010 Apr;146(4):406-411. [doi: 10.1001/archdermatol.2010.29] [Medline: 20404229]
39. Albert SM, Shevchik GJ, Paone S, Martich GD. Internet-based medical visit and diagnosis for common medical problems: experience of first user cohort. Telemed J E Health 2011 May;17(4):304-308 [FREE Full text] [doi: 10.1089/tmj.2010.0156] [Medline: 21457013]
40. Courneya PT, Palattao KJ, Gallagher JM. HealthPartners' online clinic for simple conditions delivers savings of $88 per episode and high patient approval. Health Aff (Millwood) 2013 Feb;32(2):385-392. [doi: 10.1377/hlthaff.2012.1157] [Medline: 23381532]
41. 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 Nov 01;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
42. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, Martin T, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014 Feb 01;21(e1):e157-e162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
43. Pathipati AS, Ko JM. Implementation and evaluation of Stanford Health Care direct-care teledermatology program. SAGE Open Med 2016;4:2050312116659089 [FREE Full text] [doi: 10.1177/2050312116659089] [Medline: 27493756]
44. Hawes E, Lambert E, Reid A, Tong G, Gwynne M. Implementation and evaluation of a pharmacist-led electronic visit program for diabetes and anticoagulation care in a patient-centered medical home. Am J Health Syst Pharm 2018 Jun 15;75(12):901-910. [doi: 10.2146/ajhp170174] [Medline: 29654139]
45. Levine DM, Dixon RF, Linder JA. Association of structured virtual visits for hypertension follow-up in primary care with blood pressure control and use of clinical services. J Gen Intern Med 2018 Nov;33(11):1862-1867 [FREE Full text] [doi: 10.1007/s11606-018-4375-0] [Medline: 29687432]
46. Penza KS, Murray MA, Myers JF, Maxson J, Furst JW, Pecina JL. Treating pediatric conjunctivitis without an exam: an evaluation of outcomes and antibiotic usage. J Telemed Telecare 2020;26(1-2):73-78. [doi: 10.1177/1357633X18793031] [Medline: 30153768]
47. Penza KS, Murray MA, Pecina JL, Myers JF, Furst JW. Electronic visits for minor acute illnesses: analysis of patient demographics, prescription rates, and follow-up care within an asynchronous text-based online visit. Telemed J E Health 2018 Mar;24(3):210-215. [doi: 10.1089/tmj.2017.0091] [Medline: 28731388]
48. Player M, O'Bryan E, Sederstrom E, Pinckney J, Diaz V. Electronic visits for common acute conditions: evaluation of a recently established program. Health Aff (Millwood) 2018 Dec;37(12):2024-2030. [doi: 10.1377/hlthaff.2018.05122] [Medline: 30633677]
49. Rajda J, Seraly MP, Fernandes J, Niejadlik K, Wei H, Fox K, et al. Impact of direct to consumer store-and-forward teledermatology on access to care, satisfaction, utilization, and costs in a commercial health plan population. Telemed J E Health 2018 Feb;24(2):166-169. [doi: 10.1089/tmj.2017.0078] [Medline: 28742431]
50. Murray MA, Penza KS, Myers JF, Furst JW, Pecina JL. Comparison of evisit management of urinary symptoms and urinary tract infections with standard care. Telemed J E Health 2020 May;26(5):639-644. [doi: 10.1089/tmj.2019.0044] [Medline: 31313978]
51. Penza KS, Murray MA, Myers JF, Furst JW, Pecina JL. Management of acute sinusitis via e-visit. Telemed J E Health 2020 Jun 10:epub. [doi: 10.1089/tmj.2020.0047] [Medline: 32522103]
52. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
53. Lanham HJ, Leykum LK, Pugh JA. Examining the complexity of patient-outpatient care team secure message communication: qualitative analysis. J Med Internet Res 2018 Jul 11;20(7):e218 [FREE Full text] [doi: 10.2196/jmir.9269] [Medline: 29997107]
54. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the multi-level fit between work and technology in a secure messaging implementation. AMIA Annu Symp Proc 2014;2014:954-962 [FREE Full text] [Medline: 25954403]

XSL•FO
RenderX

55. Bellon JE, Stevans JM, Cohen SM, James AE, Reynolds B, Zhang Y. Comparing advanced practice providers and physicians as providers of e-visits. Telemed J E Health 2015 Dec;21(12):1019-1026 [FREE Full text] [doi: 10.1089/tmj.2014.0248] [Medline: 26161623]
56. Tests and treatments for urinary tract infections in older people. Choosing Wisely. URL: https://www.choosingwisely.org/patient-resources/urinary-tract-infections-in-older-people/ [accessed 2021-04-21]
57. Treating sinus problems. Choosing Wisely. URL: https://www.choosingwisely.org/patient-resources/treating-sinus-problems-aaaai/ [accessed 2021-04-24]
58. Daniel SJ, Wu L, Kumar S. Teledentistry: a systematic review of clinical outcomes, utilization and costs. J Dent Hyg 2013 Dec;87(6):345-352. [Medline: 24357563]
59. Boggan JC, Shoup JP, Whited JD, Van Voorhees E, Gordon AM, Rushton S, et al. Effectiveness of acute care remote triage systems: a systematic review. J Gen Intern Med 2020 Jul;35(7):2136-2145. [doi: 10.1007/s11606-019-05585-4] [Medline: 31898116]
60. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
61. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
62. Kao DP, Lindenfeld J, Macaulay D, Birnbaum HG, Jarvis JL, Desai US, et al. Impact of a telehealth and care management program on all-cause mortality and healthcare utilization in patients with heart failure. Telemed J E Health 2016 Jan;22(1):2-11 [FREE Full text] [doi: 10.1089/tmj.2015.0007] [Medline: 26218252]
63. Grady BJ. A comparative cost analysis of an integrated military telemental health-care service. Telemed J E Health 2002;8(3):293-300. [doi: 10.1089/15305620260353180] [Medline: 12419023]
64. Russo JE, McCool RR, Davies L. VA telemedicine: an analysis of cost and time savings. Telemed J E Health 2016 Mar;22(3):209-215. [doi: 10.1089/tmj.2015.0055] [Medline: 26305666]
65. Le Goff-Pronost M, Mourgeon B, Blanchère J, Teot L, Benateau H, Dompmartin A. Real-world clinical evaluation and costs of telemedicine for chronic wound management. Int J Technol Assess Health Care 2018 Jan;34(6):567-575. [doi: 10.1017/S0266462318000685] [Medline: 30369340]
66. Nguyen HV, Tan GSW, Tapp RJ, Mital S, Ting DSW, Wong HT, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. Ophthalmology 2016 Dec;123(12):2571-2580. [doi: 10.1016/j.ophtha.2016.08.021] [Medline: 27726962]
67. Natafgi N, Shane DM, Ullrich F, MacKinney AC, Bell A, Ward MM. Using tele-emergency to avoid patient transfers in rural emergency departments: an assessment of costs and benefits. J Telemed Telecare 2018 Apr;24(3):193-201. [doi: 10.1177/1357633X17696585] [Medline: 29278984]
68. Hayward K, Han SH, Simko A, James HE, Aldana PR. Socioeconomic patient benefits of a pediatric neurosurgery telemedicine clinic. J Neurosurg Pediatr 2019 Oct 25:1-5. [doi: 10.3171/2019.8.PEDS1925] [Medline: 31653803]
69. Hale A, Haverhals LM, Manheim C, Levy C. Vet Connect: a quality improvement program to provide telehealth subspecialty care for veterans residing in VA-contracted community nursing homes. Geriatrics (Basel) 2018 Sep 05;3(3):57 [FREE Full text] [doi: 10.3390/geriatrics3030057] [Medline: 31011094]
70. Dirnberger J, Waisbren S. Efficacy of telehealth visits for postoperative care at the Minneapolis VA. Am J Surg 2020 Sep;220(3):721-724. [doi: 10.1016/j.amjsurg.2020.01.015] [Medline: 31980135]
71. Harris M. Bold consequences: the use and role of write-offs. American Association of Medical Assistants. Chicago, IL: AAMA; 2015 Jan. URL: http://www.aama-ntl.org/docs/default-source/other/fine-print-bold-consequences_article.pdf [accessed 2021-04-24]
72. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
73. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
74. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
75. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
76. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
77. Mobile fact sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-04-24]

XSL•FO
RenderX

78. Adler-Milstein J, Longhurst C. Assessment of patient use of a new approach to access health record data among 12 US health systems. JAMA Netw Open 2019 Aug 02;2(8):e199544 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9544] [Medline: 31411708]
79. Pho KK, Lu R, Gates S, Cai J, Xie D, Xie Y, et al. Mobile device applications for electronic patient portals in oncology. JCO Clin Cancer Inform 2019 Feb;3:1-8 [FREE Full text] [doi: 10.1200/CCI.18.00094] [Medline: 30802142]
80. Anderson PM, Hanna R. Defining moments: making time for virtual visits and catalyzing better cancer care. Health Commun 2020 May;35(6):787-791. [doi: 10.1080/10410236.2019.1587695] [Medline: 30907145]
81. Hernandez BF, Morgan BJ, Ish J, Agbator LO, Lindo-Moon S, Stotler FF, et al. Communication preferences and satisfaction of secure messaging among patients and providers in the military healthcare system. Mil Med 2018 Nov 01;183(11-12):e383-e390. [doi: 10.1093/milmed/usy094] [Medline: 29741659]
82. Jung C, Padman R, Shevchik G, Paone S. Who are portal users vs. early e-Visit adopters? A preliminary analysis. AMIA Annu Symp Proc 2011;2011:1070-1079 [FREE Full text] [Medline: 22195168]
83. Lee KJ, Finnane A, Soyer HP. Recent trends in teledermatology and teledermoscopy. Dermatol Pract Concept 2018 Jul;8(3):214-223 [FREE Full text] [doi: 10.5826/dpc.0803a13] [Medline: 30116667]
84. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson JA, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
85. Su MY, Das S. Expansion of asynchronous teledermatology during the COVID-19 pandemic. J Am Acad Dermatol 2020 Dec;83(6):e471-e472 [FREE Full text] [doi: 10.1016/j.jaad.2020.08.054] [Medline: 32822793]
86. Laurie H, Molly Z, Sanjana R, Megan H, Matthew H, Matthew A, et al. Expanded telehealth options during the COVID pandemic eliminated racial and age disparities in electronic health care use by IBD patients. Am J Gastroenterol 2020 Dec 01;115(Suppl 1):S20. [doi: 10.14309/01.ajg.0000723108.50072.f1] [Medline: 33566552]

Abbreviations

COVID-19: coronavirus disease 2019
e-visit: electronic communication–based visit

Edited by G Eysenbach; submitted 27.01.21; peer-reviewed by K McInnes, W Everett; comments to author 16.02.21; revised version received 02.03.21; accepted 11.04.21; published 05.05.21.

Please cite as:
Nguyen OT, Alishahi Tabriz A, Huo J, Hanna K, Shea CM, Turner K
Impact of Asynchronous Electronic Communication–Based Visits on Clinical Outcomes and Health Care Delivery: Systematic Review
J Med Internet Res 2021;23(5):e27531
URL: https://www.jmir.org/2021/5/e27531
doi:10.2196/27531
PMID:33843592

©Oliver T Nguyen, Amir Alishahi Tabriz, Jinhai Huo, Karim Hanna, Christopher M Shea, Kea Turner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2021. 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

Electronic Tools to Bridge the Language Gap in Health Care for People Who Have Migrated: Systematic Review

Frédérique Thonon[1,2], MPH, PhD; Swati Perrot[1], MSc; Abhijna Vithal Yergolkar[3], PharmD; Olivia Rousset-Torrente[1,2], MSc; James W Griffith[4], PhD; Olivier Chassany[1,2], MD, PhD; Martin Duracinsky[1,2,5], MD, PhD

[1]Patient-Reported Outcomes Unit (PROQOL), UMR 1123, Université de Paris, INSERM, F-75004, Paris, France

[2]Unité de Recherche Clinique en Economie de la Santé (URC-ECO), AP-HP, Hôpital Hôtel-Dieu, F-75004, Paris, France

[3]Faculty of Pharmacy, Ramaiah University of Applied Sciences, Karnataka, India

[4]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Service de Médecine Interne et d'Immunologie Clinique, Hôpital Bicêtre, F94270, Le Kremlin Bicêtre, France

Corresponding Author:
Frédérique Thonon, MPH, PhD
Patient-Reported Outcomes Unit (PROQOL), UMR 1123, Université de Paris, INSERM, F-75004
Hôpital Hôtel Dieu
1 place du Parvis Notre Dame
Paris, 75004
France
Phone: 33 140275543
Email: frederique.thonon@gmail.com

Abstract

Background: People who have migrated or with a language barrier may face significant hurdles in accessing health care. Some apps have been specifically developed to facilitate the dialogue between health care professionals and people who have migrated who have low-level language proficiency or to promote health among people who have migrated.

Objective: We conducted a systematic review to investigate development, acceptability, and effectiveness of these types of apps.

Methods: We conducted a search of PubMed, Scopus, and Embase databases. We included all study designs (qualitative, quantitative, mixed) reporting development, evaluation of efficacy, or acceptability of apps facilitating dialogue with a health professional or promoting health for people who have migrated, minorities, or tourists with a language barrier, using any outcome. Two researchers selected the studies independently. We collected general information about the app, information about health literacy and cultural adaptation, information about the development of the app, evidence on acceptability or efficacy, and information on app use. Data were collected by 2 researchers independently and results were reviewed to verify agreement and reported according to PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis).

Results: Positive results for translation apps included better communication, but with possible limitations, and reduced consultation time. Positive results for health promotion apps included improved quality of life and better management of chronic illnesses.

Conclusions: Overall, the apps had good levels of acceptability, though only half had their efficacy evaluated. In those evaluations, the endpoints were mostly related to reported behavior change and knowledge improvement, which is common for evaluations of health promotion programs. In the future, as more health apps are created, it is essential that apps that claim to have a public health objective undergo a rigorous evaluation of their acceptability, efficacy, and actual use. Indicators of outcomes beyond changes in behavior and knowledge should be reported; change in health status or access to care should also be reported. This systematic review has helped us note the characteristics associated with improved acceptability and efficacy, which can be helpful for the development of future apps.

(J Med Internet Res 2021;23(5):e25131) doi:10.2196/25131

KEYWORDS

eHealth; systematic review; migrants; health literacy; access to care; health promotion

XSL•FO
RenderX

Introduction

People who have migrated may face significant delays and barriers in accessing health care, especially those who do not fluently speak the language of the host country. Research has been conducted to investigate language barriers in accessing care and prevention among people who have migrated and its health consequences. Pregnant women who have migrated and with low proficiency in the language of their new country of residence have found it more difficult to access care [1], had less knowledge about the benefits of folic acid and had lower folic acid intake [2], and had a higher risk of obstetric trauma [3]. A study [4] showed that a language barrier is significantly associated with a higher occurrence of serious medical events in pediatrics. Cancer studies show less screening for colorectal cancer [5,6], cervical cancer [6], and breast cancer [6,7] among people who have migrated who had low English proficiency. Others have found that adding an interpreter to a consultation where the patient has a language barrier (any type of interpretation) results in the reduction of obstetric interventions [8], better clinical outcomes in people with diabetes [7,8], a higher rate of breast cancer and colorectal cancer screening [7], and a higher rate of influenza vaccination [7]. Overall, it has also been shown that patients with a low proficiency in the language of the country of residence receive more preventive advice and more prescriptions and have fewer emergency visits if they use interpretation services [9].

Health care professionals commonly use several solutions to communicate with people who have migrated with low language proficiency, including printed guides and brochures, informal interpreters, professional interpreters in person, professional interpreters on phone or video conference, and general translation apps. Some apps or electronic tools have been specifically developed to be used in medical consultations to facilitate dialogue between health care professionals and people who have migrated who have low language proficiency or to promote health among people who have migrated. In the last decade, there has been a sharp increase in the number of medical apps developed [10], but their impact on patients or public health should be established.

We conducted a systematic review to examine the evidence related to the development, adaptation, acceptability, and effectiveness of electronic tools designed to help health care providers communicate with or promote health among people who have migrated and who have low levels of proficiency in the language of their country or low levels of health literacy. The aim of this review was to describe the existing tools and gather evidence about features that increase the acceptability and efficacy of such tools. Our work is part of a larger project designed to develop and evaluate an app to facilitate communication between people who have migrated who face a language barrier and health professionals regarding testing for HIV and different forms of viral hepatitis [11]. Lessons learned and evidence from similar apps could help design an app that has a significant impact on public health.

Methods

Search Strategy, Databases, and Keywords

We wrote a protocol prior to starting the systematic review and reported the results according to the PRISMA reporting guideline [12] (Multimedia Appendix 1). We conducted a search of 3 databases of scientific publications: PubMed, Scopus, and Embase. The keywords varied according to the database used. To determine keywords, we undertook a broad preliminary search and selected some articles (between 5 and 10) identified as meeting inclusion criteria determined by the authors. After selecting these articles, we searched for their related keywords, MeSH terms (Medical Subject Headings), terms in titles or abstracts, and EmTree terms. Then the selection of keywords was tested and different combinations were tested so that the total number of results was manageable and yielded relevant articles. The final selection of keywords was critically reviewed by a university librarian. The keywords for each database are detailed in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

We used the inclusion and exclusion criteria detailed in Table 1. Although the primary focus of this review was people who have migrated, the preliminary search yielded articles reporting apps that helped bridge the language barrier for other populations, such as tourists needing emergency care and not speaking the language of the country visited or indigenous people whose primary language is different from the official language. Since these apps help health care providers communicate with patients with a language barrier, we decided to include them because we found they were relevant.

XSL•FO
RenderX

Table 1. Inclusion and exclusion criteria.

Criteria type	Inclusion criteria	Exclusion criteria
Publication		
Language	• Written in English or French	• Other languages
Date range	• Published after 1998	• Published before 1998
Type	• Original article, review, protocol, conference abstract, book chapter	• Other types of publications: editorial, letter, notes, etc
Study		
Design	• Studies reporting the development of an electronic tool, including qualitative or quantitative studies of people who have migrated or health providers, mixed methods, literature reviews • Studies evaluating the acceptability of electronic tools, including qualitative or quantitative studies, usability studies, randomized or nonrandomized trials • Studies evaluating the efficacy of electronic tools, including randomized or nonrandomized trials, qualitative or quantitative studies, economic evaluations	• Articles lacking information about the development or evaluation of an electronic tool • Studies exploring only the perceptions of users (people who have migrated or health professionals) related to e-health or a health issue
Population		
Language	• International people who have migrated not fluent with the language of the country they reside	• People with no language barrier (internal people who have migrated, ethnic minorities, people who have migrated with no language barrier)
Communication barriers	• Cultural minorities having a language barrier (eg, indigenous people whose first language is different from the official language) • Tourists	• People with other type of communication barriers: deaf or hard-hearing people, people with a learning disability
Technology		
	• Website, mobile (smartphone or tablet) apps, other electronic technology that allows interaction with user text message or email-based services	• Tools using only print material, audio, or video
Intervention	• Technology designed to help communication between health care providers (eg, doctors, nurses, midwifes) and people who have migrated in any health care setting (eg, hospital, primary care) • Technology designed to promote healthy behavior among people who have migrated	• Technology that aims to facilitate communication or translation in general settings but not designed specifically for the medical setting (eg, Google Translate, apps for tourists).
Outcome		
Development of an electronic tool	• Themes emerging from interviews or focus groups, results from participants consultations	• None
Acceptability	• Comments from participants,, satisfaction surveys, data in app use or consultations	• None
Efficacy	• Changes in health outcomes (self-reported or measured with biomedical measures), changes in knowledge, attitudes, practices, and beliefs	• None

Study Selection

A list of articles retrieved from all 3 journal databases was compiled. After excluding duplicates, 2 researchers (FT and SP) independently reviewed the title and abstracts of all documents for preselection. Articles were then reviewed in full for inclusion or exclusion; if an article was excluded, the reason was documented. We managed the selection of articles with Rayyan [13]. Differences of opinion regarding the inclusion of an article were managed by a third researcher (ORT). We also

XSL•FO
RenderX

subsequently included relevant articles that were cited by articles that had been initially selected.

Data Collection

We used 2 data collection coding sheets (Google Forms): one for each article (type of publication, year of publication, and journal) and one for each app studied (general information about the app or electronic tool, information about health literacy and cultural adaptation, information about the development of the app or electronic tool, evidence about the acceptability or efficacy of the app or electronic tool, information about the use of the app or electronic tool). All articles were read by FT, and to improve the validity of results, we performed data triangulation by having a second author read articles independently.

When articles presented additional sources of information regarding an app, such as gray literature or a website, we retrieved data about the app from this source and noted the references of this additional source of information but did not use them in the article selection.

Quality and Risk of Bias Assessment

We evaluated the articles using ICROMS (Integrated Quality Criteria for Review of Multiple Study Designs) [14], which can be used for public health reviews that include several study designs, such as randomized controlled trials, controlled before-after, controlled interrupted times series, cluster randomized controlled trial, noncontrolled before-after, cohort studies and qualitative studies. This tool consists of 33 indicators grouped in 7 dimensions: (1) clear aims and justification; (2) managing bias in sampling or between groups, (3) managing bias in outcome measurement and blinding, (4) managing bias in follow-ups, (5) managing bias in other study aspects; (6) analytical rigor; and (7) managing bias in reporting or ethical considerations. Each indicator receives a score of 2 if the criteria for the indicator are met, 0 if this is not the case, or 1 if it is unknown whether the criteria were met. Where specified, we also found that it was necessary to use CHEERS (Consolidated Health Economic Evaluation Reporting Standards [15]), noting if the information required was available, incomplete, or not available using the same scoring system as ICROMS. It should be noted that this scoring system for protocols or medicoeconomic studies is not validated, nor is it part of the ICROMS tool.

Many articles related to eHealth are information technology usability studies. Usability studies are studies that aim to explore usability requirements, discover usability problems, and design solutions [16]. To our knowledge, there is no published article proposing a tool to evaluate the quality or bias of such studies. Usability studies mainly use either qualitative or quantitative methodologies [16], however, due to the usually small sample of participants [17] and approaches used, such as participants being asked to perform tasks and give oral feedback [18], they are closer to qualitative studies than they are to quantitative studies. Therefore, we assessed the quality and risk of bias of most of usability studies using the ICROMS tool adapted to qualitative studies, except usability studies using quantitative-only methodology, which were assessed as noncontrolled before-after. We also used the ICROMS tool to evaluate research protocols; we simply disregarded the questions that were not applicable.

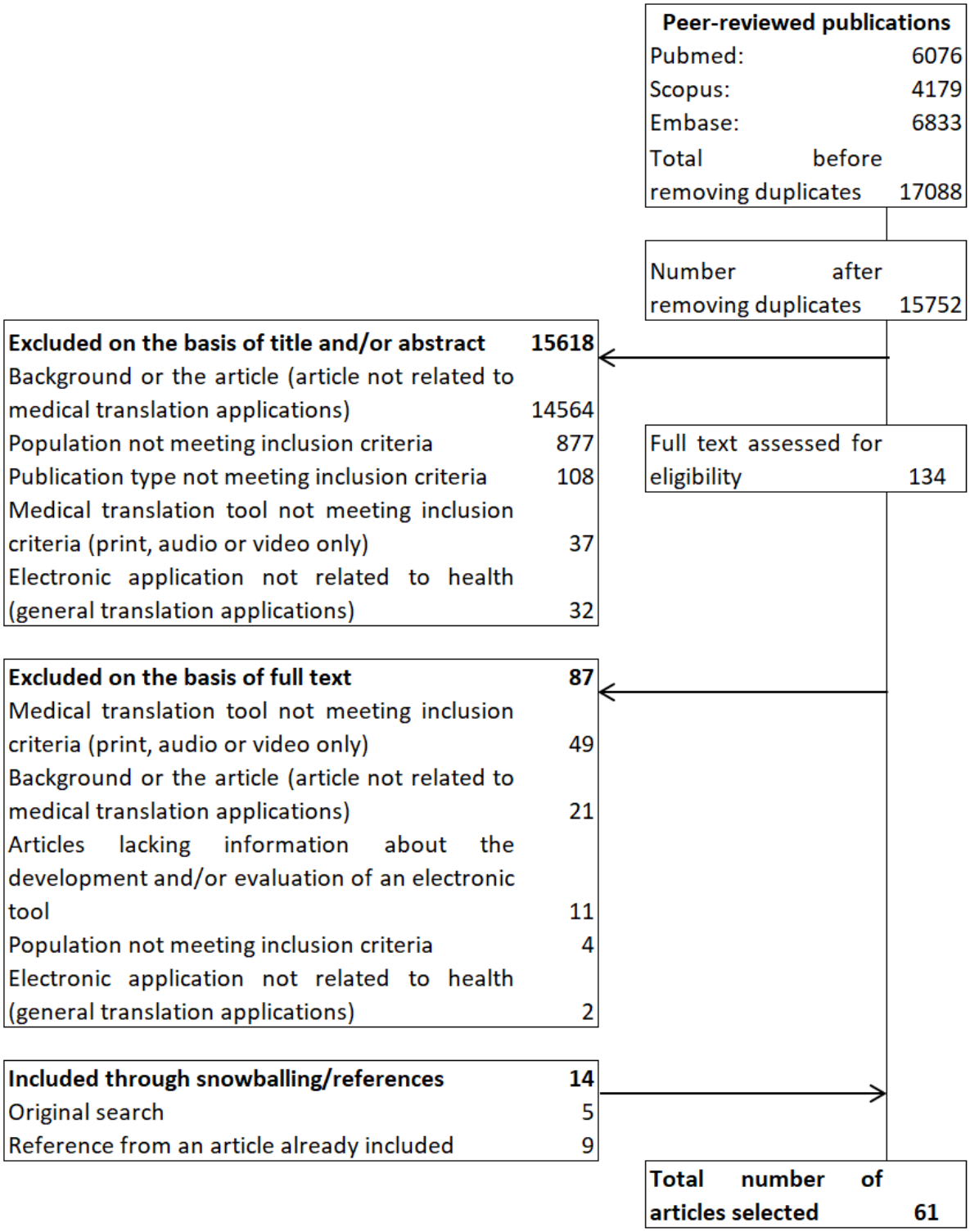
Results

Selection

The database search was carried out in October 2019. We retrieved a total of 15,752 articles from 3 databases after removing duplicates, of which 15,618 were excluded on the basis of title or abstract content. We assessed the full texts of 134 articles and excluded 87, most because the electronic tool described did not meet the inclusion criteria. We subsequently added 14 articles, because they were found either in the original background search or cited by articles that had been initially selected. In total, we included 61 articles (Figure 1). The difference between the very high number of articles originally retrieved and the number of articles selected can be explained by the fact that the original search yielded a number of articles related to translational research. The 61 selected articles were all read and analyzed by FT, and 51 (84%) were read and analyzed by a second reviewer (35 by SP and 16 by AVY).

XSL•FO
RenderX

Figure 1. Flowchart of article selection.



Characteristics

The majority of articles were original research articles (55/61, 90%); 4 were published research protocols, and 2 were proceedings of scientific conferences. Of the 61 articles, most articles (27/61, 44%) were published in journals that specialized in medical information technology, 14 (23%) were published in journals related to a disease or medical specialty, 11 (18%) were published in journals that specialized in health promotion, 6 (10%) were published in general health journals, and 3 (5%) were published in nonmedical journals that specialized in information technology. All articles were published in English. The countries of affiliation of the first authors were USA (n=40), Germany (n=5), Australia (n=5), New Zealand (n=2), United Kingdom (n=2), Italy (n=1), Nigeria (n=1), China (n=1), Switzerland (n=1), Japan (n=1), Norway (n=1), and Spain (n=1).

Of the 61 articles, 21 articles reported qualitative studies (34%), 18 articles reported usability studies (30%), 9 articles reported

XSL•FO
RenderX

nonrandomized interventional studies (15%), 7 articles reported randomized controlled trials (11%), 4 articles were protocols for either randomized controlled trials (n=3, 5%) or a mixed methods study (n=1, 2%), and 2 articles reported other types of studies (1 medicoeconomic study; 1 case study).

Quality and Risk of Bias

One article was a health economic study, for which CHEERS was used. We could not evaluate the quality of one article [19], as it is was a description of the tool and a single case study of its use on a patient. This article was included in our review because of its practical description of the tool.

The full results are available in Multimedia Appendix 3. Among the 7 randomized controlled trial articles, the mean score was 22 and the median score was 24 (the score ranges from 0 to 30, and the minimum score required for a study to be considered of robust quality is 22); 1 article did not meet the criterion to be considered of robust quality. Among the 21 qualitative studies, the average and median score were 18 and 18, respectively (the score ranges from 0 to 26, and the minimum required for a study to be considered of robust quality is 16); 7 articles did not meet the minimum to be considered of robust quality. Among the 9 noncontrolled before-after studies, the mean score was 22 and the median score was 23 (the score ranges from 0 to 30 and the minimum required for a study to be considered of robust quality is 22); 4 studies did not meet the minimum to be considered of robust quality. Among the 18 usability studies, 6 did not meet the minimum required for either qualitative or noncontrolled before-after studies to be considered of robust quality.

For the randomized controlled trials, the lowest scores were found on questions regarding allocation blinding and measurement blinding, and to a lower extent, reliability of primary outcome measures (lack of objectivity of outcome variables). For the qualitative studies and usability studies, the lowest scores were found for items of critically assessing researcher bias and lack of agreement. For noncontrolled before-after studies, the lowest scores were found regarding justification for and attempts to mitigate the lack of control group and, to a lower extent, lack of objectivity of outcome measures.

General Characteristics of Apps

In the 61 articles, a total of 48 apps were presented (Table 2). Approximately two-thirds of the electronic tools (n=30, 63%) were developed in the USA. Other countries represented were Germany (4 apps), Australia (4 apps); New Zealand (2 apps) and Italy, United Kingdom, Nigeria, China, Switzerland, Japan, Norway, and Spain (1 app each).

Of the 48 apps, 20 apps (42%) were designed for health promotion or prevention, rather than for one specific health care setting: 11 were designed for hospital care, 8 were designed for primary care, 5 were designed for therapeutic or patient education, 3 were designed for both primary care and hospital care, and one was designed for both health promotion and primary care; 14 apps (29%) were health promotion apps that were not related to one specific medical specialty or condition, and the most represented medical specialties were cancer (9/48, 19%), mental health (6/48, 13%). Health promotion is defined as the process of enabling people to increase control over and to improve their health [20].

One-third of apps (16/48, 33%) were designed solely to facilitate interactions between people who have migrated and health care providers during a consultation, while the remaining two-thirds (32/48, 67%) were designed to promote health among people who have migrated who face a language barrier. Of those 32 apps designed to promote health, 9 were adaptations for people who have migrated of existing apps, while 23 were new apps developed specifically for people who have migrated.

Most electronic tools (38/48, 79%) were in the form of a mobile app, while other types (text messaging, website) were less common. It should be noted that 2 electronic tools (2/48, 4%) included both a mobile app and a text messaging service. Almost two-thirds of apps (31/48, 65%) were interactive, meaning that they allowed feedback from the user, either to another user or to the app.

More than three-quarters (38/48, 79%) were specifically targeted to a group of people who have migrated, and mostly for people who have migrated from specific nationalities (24/38, 63%). Over two-thirds (33/48; 69%) had 1 language in addition to the source language, nine apps had 3 to 9 languages, five had 10 to 19 languages, and one app had over 20 languages. Information about the funding of the app was available for 33 (69%) out of the 48 apps: 30 apps (63%) had received a funding from either a public source, charitable source or crowd-funding, while 2 apps (4%) had received a funding from a mix of private and public or charitable source, and 1 app (2%) had received a private or industry funding.

XSL•FO
RenderX

Table 2. Characteristics of apps included in the analysis.

Characteristic	Apps (n=48), n (%)
Country of development	
USA	30 (63)
Germany	4 (8)
Australia	4 (8)
New Zealand	2 (4)
Italy	1 (2)
United Kingdom	1 (2)
Nigeria	1 (2)
China	1 (2)
Switzerland	1 (2)
Japan	1 (2)
Norway	1 (2)
Spain	1 (2)
Setting	
Health promotion/prevention	20 (42)
Hospital care	11 (23)
Primary care	8 (17)
Therapeutic or patient education	5 (10)
Primary care and hospital care	3 (6)
Health promotion/prevention and primary care	1 (2)
Medical specialty	
Health promotion without a focus on a medical specialty	14 (29)
Cancer	9 (19)
Mental health, psychiatry	6 (13)
Infectious diseases	5 (10)
Cardiovascular diseases, endocrinology (diabetes)	3 (6)
Gynecology, pregnancy	3 (6)
Emergency medicine, intensive care	2 (4)
Addiction medicine	2 (4)
Paramedical specialties (physiotherapy, occupational therapy, speech pathology, dietetic, podiatrists)	2 (4)
Pediatrics	1 (2)
Pulmonology	1 (2)
Aim	
Facilitating communication between migrant and health provider	16 (33)
Promoting healthy behavior among people who have migrated[a]	32 (67)
Including adaptation of existing apps	9
Features	
Mobile app	38 (79)
Text-messaging service	6 (13)
Mobile app and text-messaging service	2 (4)
Website for consultation	2 (4)
Interactivity	

XSL•FO
RenderX

Characteristic	Apps (n=48), n (%)
Interactive	31 (65)
Not interactive	17 (35)
Target population	
People who have migrated from specific nationalities	24 (50)
Asylum seekers/refugees	4 (8)
People who have migrated at risk from their occupation	3 (6)
Indigenous people	3 (6)
People who have migrated who are concerned by a specific health condition	3 (6)
Other	1 (2)
The app does not specifically target a group of people who have migrated	10 (21)
Languages (in addition to the source language)	
1	33 (69)
2-9	9 (19)
10-19	5 (10)
≥20	1 (2)
Fee required	
Yes	5 (10)
No	9 (19)
Don't know	34 (71)
Institution funding the development	
Public/government	11 (23)
Charitable or crowdfunding	3 (6)
Mix of charitable and public/government	2 (4)
Public/government, private and charitable	1 (2)
Private/industry only	15 (31)
No information about the funding of the app	16 (33)

[a]Of the 32, 9 were adaptations of existing apps.

Health Literacy and Cultural Adaptation

There was information about how the translation was performed for only half of the apps (n=24); translation was performed by a professional translator (16/48, 33%); informally (4/48, 8%), usually by nonprofessional native speakers; or by a mix of professional and personal translation (4/48, 8%). Some form of quality control of the app was mentioned for 18 apps (38%), usually forward-backward translation or translation being checked by a native speaker. The translation included cultural adaptation on 23 apps (48%), half of the apps included pictures or pictograms (n=24). Two-thirds of apps had either an audio or video feature (32/48, 67%).

Information About the Development of the App or Electronic Tool

Of the 48 apps, 33 (69%) reported using scientific methods to develop the content of the app: qualitative studies (interviews or focus groups) for 15 (45%), use of a theoretical framework such as behavioral theories for 8 (24%), use of existing guidelines of curriculums (n=5, 15%), a mix of qualitative and quantitative methods for 4 (12%), and a mix of survey and use of theoretical framework for 1 (3%).

Of the 48 apps, less than half (22/48, 46%) reported involving users (people who have migrated) in the development, 18 apps (38%) reported involving health professionals in the development, and 5 apps (10%) reported involving other stakeholders such as charities.

Evidence About the Acceptability and Efficacy of the App or Electronic Tool

Many apps (32/48, 67%) had their acceptability evaluated (translation apps: 14/16, 88%; health promotion apps: 18/32, 56%). Of the 14 translation apps that had their acceptability evaluated, 3 were evaluated among people who had migrated only, 6 were evaluated among people who had migrated and health professionals, and 5 were evaluated among health professionals alone. Of the 18 apps designed to promote health among people who had migrated that have been evaluated for acceptability, 14 were evaluated among people who had migrated alone, 3 were evaluated among people who had

XSL•FO
RenderX

migrated and health professionals, and 1 was evaluated among health professionals alone. Acceptability studies used mixed methods (20/32, 63%), used quantitative methods only (8/32, 25%), used qualitative methods only (7/32, 22%), were pilots (5/32, 16%), and randomized controlled trials (2/32, 6%). The endpoints used to measure acceptability were comments from participants (12/32, 38%), results from satisfaction survey (9/32, 28%), length of time of a consultation and satisfaction survey (4/32, 13%), data on the use of the app (3/32, 9%), a mix of survey and qualitative comments (2/32, 6%), and other outcomes (2/32, 6%). The Systems Usability Scale was mentioned in 5 evaluations. Other evaluation systems mentioned were the Technology Acceptance Model and the Stanford Communication with Physicians Scale. Among the 32 apps that had their acceptability evaluated, 25 (78%) reported an overall good or very good acceptability; 1 (3%) reported an adequate acceptability; for 3 apps (9%), the study was ongoing; and for 3 others, the results cannot be reported as they consisted of comments from participants or choice of a design. Of the 25 apps that reported a good or very acceptability, 8 had been developed involving users in the process.

Half of the apps (n=24) had their efficacy evaluated (translation apps: 2/16, 13%; health promotion apps: 22/32, 69%). Study designs were randomized controlled trials (9/24, 38%), nonrandomized trials (7/24, 29%), survey (3/24, 13%), qualitative (2/24, 8%), mixed methods (2/24, 8%), and economic analysis (1/24, 4%). The endpoints or outcomes used for those evaluations were reported behavior change (10/24, 42%), knowledge improvement (7/24, 29%), self-reported health markers, such as improved quality of life, better sleep, less anxiety (4/24, 17%), biometric health markers (2/24, 8%), cost-effectiveness (2/24, 8%), and accuracy of medical information (n=2, 8%). The total exceeds 100% as 11 apps used several different endpoints. Among the 24 apps that had their efficacy evaluated, 12 (50%) had significant positive results; 5 (12%) had partially positive results, meaning that the app showed significant efficacy in some measured outcomes but not all, or was effective in some population and not all; and 2 apps had nonsignificant results (8%); 5 studies were ongoing (21%). Details of the efficacy and acceptability studies are in Multimedia Appendix 4.

Positive outcomes (Table 3) reported for translation apps includes reducing the need to call an interpreter, especially in emergency situations [21], reduced consultation time [22,23], and reduced patient anxiety [24]. Negatives that were reported included limitations in the dialogue between health professionals and patients [25-27] and concerns about hindering the therapeutic relationship [28]. Health promotion apps had positive results in terms of acceptability and efficacy. Positive outcomes included improved quality of life and better management of chronic illnesses such as diabetes [29-31], cancer [32-34], HIV [35], depression [19,36], and addiction [37,38].

Table 3. Characteristics of health apps linked to better acceptability or efficacy.

App type	Important points noted during the development	Characteristics linked to better acceptability or efficacy
Translation	Many experts recommend that culturally tailored materials be created de novo or in tandem, rather than as variations on existing materials	• Speech is generally preferred to text • Including a button, equivalent to the patient's "I do not understand the question" • Including a phrase for health care practitioners "I don't understand your answer" • Integrating an option to directly call an interpreter in the app • Integrating a list of nearby hospitals for follow-up care • Including the option for patients to respond with pictures • Including the option for health care practitioners to save the conversation with a patient (with respect to data protection and confidentiality)
Health promotion	Addressing both motivation as well as linguistic and sociocultural barriers and reassuring participants of confidentiality	• Apps that personalize the experience for users are preferred • Including a help function and a tutorial • Include a tutorial provided by a virtual human rather than text • Culturally appropriate, with photos of a multigenerational family • Colorful and eye catching, but also professional, with easy-to-access information • Easy to navigate with simple and easy-to-understand information • Interactive with immediate feedback • Audio, videos and pictures • Provision of links for further information about a health issue • Use of humor considered very effective by target audiences • Including a list of frequently asked questions for users • Including the option for people who have migrated to learn medical terms in the language of the country they live in

Discussion

Main Results

We synthesized evidence regarding the development, acceptability, and efficacy of health apps and electronic tools created to overcome the language barriers. Acceptability was evaluated in almost two-thirds of the apps and was generally high. Although health promotion or prevention programs specifically targeting people who have migrated might be complex to integrate into a health system, they are generally well accepted [39] and are a way to elevate health for all individuals.

Efficacy evaluations were only conducted for half of the apps. In those evaluations, the endpoints were mostly related to

XSL•FO
RenderX

reported behavior change and knowledge improvement. Knowledge improvement, however, does not systematically lead to behavior change and the reported behavior change may not be long lasting. Changes in health outcomes are rarely measured in health promotion programs or health promotion research; this should be included as a goal of health communication tools. Indeed, a survey of researchers in health promotion highlighted that the majority of assessment measures changes in awareness, knowledge, skills, policy changes, changes in behavior, changes in community capacity, changes in organizational capacity, but changes in health outcomes are not cited [40]. In addition, several systematic reviews [41-43] examining the literature related to health promotion programs have shown that programs are rarely evaluated in terms of health outcomes. It seems, therefore, that assessments of health promotion apps are in line with current practices for assessments of health promotion programs, which focus on knowledge and reported practices but rarely on final health outcomes. Given the difficulty of measuring the health outcomes of health promotion programs or apps, it is necessary to develop new methods.

Although we could only retrieve information about the funding of 69% of apps (33/48), we found that the majority had either public or charitable funding, while only 2 apps received industry funding. Additionally, only 5 apps reported charging users a fee. That seems to suggest that both health promotion and medical translation apps mostly have a noncommercial purpose and were designed with a public health goal.

Only half of the apps had their efficacy evaluated, of which half had a significant positive result. Efficacy was evaluated more often for health promotion apps than for medical translation apps. As we have mentioned, communication difficulties between health providers and individuals with language barriers can have several negative consequences, such as less satisfaction with care [7], longer consultation time [44], lower adherence to treatment protocols [45], less health education messages delivered [9], and worse clinical outcomes [8]; therefore, it is of tremendous importance that medical translation apps are rigorously assessed, not only for their acceptability but also for their efficacy.

Comparison With Literature

Several literature reviews [46-48] have been conducted of health apps and multimedia-based health promotion programs; however, none has specifically examined apps that are focused specifically on language barriers. Two reviewed the use of mobile health technology use and implications in historically underserved and minority populations in the United States [46] and mobile health interventions to promote physical activity for Black and Hispanic women [47] but scarcely addressed language barriers. We found one systematic review evaluating consumer health information technology interventions toward US Spanish-speaking populations [48]. The study [48] focused on one specific population in a specific country (USA), for a very wide scope of electronic interventions (eg, radio, videos, text messages services) that is different from the scope of our article. In that systematic review, the most commonly used evaluation metrics were behavior, attitude change, usability, and knowledge retention. The results of the study [48] and of our own were similar.

Strengths and Limitations

We conducted an extensive search of the literature using 3 different databases of publications. The selection of articles was conducted by two researchers working independently; results were compared and differences were settled by a third party. Most articles were read and their data extracted by two researchers independently, and results were compared when analyzing data. This enabled us to have a higher quality of data and a reduced risk of bias. We extended the scope of our systematic review to both electronic tools designed for translation and those designed to promote health among people who have migrated with a language barrier. Although the electronic tools have seemingly different objectives, it is likely that, in the future, there will be hybrid apps that will be developed to integrate both objectives. Indeed, many primary care consultations include health education and advice from a health provider, and hospitals have been advised to include health promotion activities into their activities [49]. Development of new hybrid apps that includes both objectives will be able to learn from the evidence from the both types of apps we examined.

In this systematic review, we only included apps that were referred to in a scientific journal. Our analysis did not include the plethora of apps designed to promote heath or facilitate consultation for people who have migrated with a language barrier that were not the topic of published peer-reviewed articles. Since the objective of our review was to evaluate the evidence related to the development or evaluation of these types of apps, it was not relevant to include these other sources. In our review, we did not assess the technical characteristics of the apps studied. Indeed most articles included in this review did not give information about the technical characteristics of the apps. Many scales and evaluations systems have been created to that end [50-53]. Assessing the technical characteristics is time-consuming and not always possible as many of those apps are not available for public use. We suggest that when new apps are developed, they strive to achieve the technical qualities measured by such scales. Other systematic reviews [54] have examined an extensive range of apps designed to help communication between health care providers and people who have migrated beyond those published in peer-reviewed articles, which focus on technical characteristics, and which provide different types of information that are complementary to that provided herein.

Implication for Policy and Conclusion

As previously mentioned, people who do not speak the language of the country have poorer access to care, longer and less satisfactory consultations, and worse clinical outcomes. In this review, we found that translation apps showed good user satisfaction but had less data on changes in the process of care (consultation length, renouncing medical care) and no data on possible changes in clinical outcomes or medicoeconomic benefits. The evaluations of health promotion apps had positive results in terms of acceptability and efficacy; however the trials on efficacy mostly used self-reported outcomes, such as

XSL•FO
RenderX

self-reported behavior changes or quality of life, rather than clinical outcomes. Most trials lacked randomization or control groups, blinding, or objective measures. This lower quality makes it difficult to draw clear conclusions on efficacy.

Future apps that are developed should include evaluation of clinical and possibly medicoeconomic benefits to draw clear recommendations on their use. The apps that were the most acceptable were those that integrated features beyond simple translation, such as making appointments with health professionals on the platform or entering basic information to prepare a visit. We recommend that future translation apps are created for medical visits integrate such. In both translation and health promotion apps, including audio and video features was most appreciated by users. We recommend integrating such features in the development of new apps.

In the future, more and more health apps will be created. Given the high cost of development [55], it is essential that apps that claim to have a public health objective undergo a rigorous evaluation of their acceptability, and efficacy. Future studies should also use strong epidemiological indicators as outcomes, such as changes in health status or access to care, rather than only using reported changes in behavior and knowledge.

The aim of this systematic review was to examine the evidence related to the development, adaptation, acceptability, and effectiveness of electronic tools designed to help health care providers communicate with or promote health among people who have migrated having a low proficiency in the language of the country of origin or a low level of health literacy. Our results, especially development and characteristics associated with a better acceptability and efficacy, should be of help to public health professionals who develop new apps.

Acknowledgments

We sincerely thank Corinne H Miller, Clinical Informationist at Galter Health Sciences Library & Learning Center of Northwestern University, Chicago, USA, for her help in the developing the search strategy and keywords.

The study was funded by the *Agence Nationale de Recherches sur le SIDA et les Hépatites virales* (grant ECTZ109287) and by *Office Français de l'Immigration et de l'Intégration* (20191104122918510), whom we thank for their support. The funders had no role in study design, data collection or analysis, manuscript writing, or decision to publish. JWG was supported in part by the National Institutes of Health (R01MD010440).

Authors' Contributions

FT, ORT, MD, SP, JWG, and OC designed the study. FT and SP developed the search and selection strategy, developed the data extraction form, and collected data. AVY collected data. FT, ORT, MD, and OC performed critical evaluation and discussion of results. FT drafted the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA reporting checklist completed.
[DOCX File , 20 KB - jmir_v23i5e25131_app1.docx]

Multimedia Appendix 2
Keywords used for the database search.
[DOCX File , 15 KB - jmir_v23i5e25131_app2.docx]

Multimedia Appendix 3
Studies quality and risk assessment.
[DOCX File , 64 KB - jmir_v23i5e25131_app3.docx]

Multimedia Appendix 4
Summary of apps studied in this systematic review.
[DOCX File , 71 KB - jmir_v23i5e25131_app4.docx]

References

1. Hoang H, Le Q, Kilpatrick S. Having a baby in the new land: a qualitative exploration of the experiences of Asian migrants in rural Tasmania, Australia. Rural Remote Health 2009 Mar 31;9(1):1-11. [doi: 10.22605/rrh1084]
2. van Eijsden M, van der Wal MF, Bonsel GJ. Folic acid knowledge and use in a multi-ethnic pregnancy cohort: the role of language proficiency. BJOG 2006 Dec;113(12):1446-1451. [doi: 10.1111/j.1471-0528.2006.01096.x] [Medline: 17081188]

XSL•FO
RenderX

3. Sentell T, Chang A, Ahn HJ, Miyamura J. Maternal language and adverse birth outcomes in a statewide analysis. Women Health 2016 Sep 11;56(3):257-280 [FREE Full text] [doi: 10.1080/03630242.2015.1088114] [Medline: 26361937]
4. Cohen AL, Rivara F, Marcuse EK, McPhillips H, Davis R. Are language barriers associated with serious medical events in hospitalized pediatric patients? Pediatrics 2005 Sep;116(3):575-579. [doi: 10.1542/peds.2005-0521] [Medline: 16140695]
5. Sentell T, Braun KL, Davis J, Davis T. Colorectal cancer screening: low health literacy and limited English proficiency among Asians and Whites in California. J Health Commun 2013;18 Suppl 1:242-255 [FREE Full text] [doi: 10.1080/10810730.2013.825669] [Medline: 24093359]
6. Thompson CA, Gomez SL, Chan A, Chan JK, McClellan SR, Chung S, et al. Patient and provider characteristics associated with colorectal, breast, and cervical cancer screening among Asian Americans. Cancer Epidemiol Biomarkers Prev 2014 Nov 02;23(11):2208-2217. [doi: 10.1158/1055-9965.epi-14-0487]
7. Flores G. The impact of medical interpreter services on the quality of health care: a systematic review. Med Care Res Rev 2005 Jun 18;62(3):255-299. [doi: 10.1177/1077558705275416] [Medline: 15894705]
8. Karliner LS, Jacobs EA, Chen AH, Mutha S. Do professional interpreters improve clinical care for patients with limited English proficiency? a systematic review of the literature. Health Serv Res 2007 Apr;42(2):727-754 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00629.x] [Medline: 17362215]
9. Jacobs EA, Shepard DS, Suaya JA, Stone E. Overcoming language barriers in health care: costs and benefits of interpreter services. Am J Public Health 2004 May;94(5):866-869. [doi: 10.2105/ajph.94.5.866] [Medline: 15117713]
10. Sweileh WM, Al-Jabi SW, AbuTaha AS, Zyoud SH, Anayah FMA, Sawalha AF. Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016. BMC Med Inform Decis Mak 2017 May 30;17(1):72 [FREE Full text] [doi: 10.1186/s12911-017-0476-7] [Medline: 28558687]
11. Thonon F, Fahmi S, Rousset-Torrente O, Bessonneau P, Griffith J, Brown C, et al. Promoting HIV, Hepatitis B Virus, and Hepatitis C Virus Screening Among Migrants With a Language Barrier: Protocol for the Development and Evaluation of an Electronic App (Apidé). JMIR Res Protoc 2021 May 5;10(5):e22239 [FREE Full text] [doi: 10.2196/22239] [Medline: 33949963]
12. 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 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
13. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
14. Zingg W, Castro-Sanchez E, Secci F, Edwards R, Drumright L, Sevdalis N, et al. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). Public Health 2016 Apr;133:19-37. [doi: 10.1016/j.puhe.2015.10.012]
15. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health 2013 Mar;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
16. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 May;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
17. Lewis JR. Sample sizes for usability studies: additional considerations. Hum Factors 1994 Jun;36(2):368-378. [doi: 10.1177/001872089403600215] [Medline: 8070799]
18. Nielsen J, Clemmensen T, Yssing C. Getting access to what goes on in people?s heads? reflections on the think-aloud technique. In: Proceedings of the second Nordic conference on Human-computer interaction. 2002 Presented at: Second Nordic conference on Human-Computer Interaction; October 19-23; Aarhus, Denmark p. 101. [doi: 10.1145/572020.572033]
19. Dahne J. ¿Aptívate!: a Spanish-language behavioral activation mobile app for delivery via primary care. Psychol Serv 2019;16(2):275.
20. Ottawa Charter for Health Promotion. World Health Organization. 1986. URL: https://www.euro.who.int/fr/publications/policy-documents/ottawa-charter-for-health-promotion,-1986 [accessed 1986-11-17]
21. Day KJ, Song N. Attitudes and concerns of doctors and nurses about using a translation application for in-hospital brief interactions with Korean patients. J Innov Health Inform 2017 Oct 06;24(3):916 [FREE Full text] [doi: 10.14236/jhi.v24i3.916] [Medline: 29121849]
22. Silvera-Tawil D, Pocock C, Bradford D, Donnell A, Harrap K, Freyne J, et al. CALD Assist-Nursing: improving communication in the absence of interpreters. J Clin Nurs 2018 Nov 30;27(21-22):4168-4178. [doi: 10.1111/jocn.14604] [Medline: 29968388]
23. Freyne J, Bradford D, Pocock C, Silvera-Tawil D, Harrap K, Brinkmann S. Developing digital facilitation of assessments in the absence of an interpreter: participatory design and feasibility evaluation with allied health groups. JMIR Form Res 2018 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/formative.8032] [Medline: 30684405]

XSL•FO
RenderX

24. Taicher BM, Alam RI, Berman J, Epstein RH. Design, implementation, and evaluation of a computerized system to communicate with patients with limited native language proficiency in the perioperative period. Anesth Analg 2011;112(1):106-112. [doi: 10.1213/ane.0b013e3182009cd1]
25. Somers HL, Caress A, Evans DG, Johnson MJ, Lovel HJ, Mohamed Z. A computer-based aid for communication between patients with limited English and their clinicians, using symbols and digitised speech. Int J Med Inform 2008 Aug;77(8):507-517. [doi: 10.1016/j.ijmedinf.2007.10.007] [Medline: 18061536]
26. Spechbach H, Gerlach J, Mazouri Karker S, Tsourakis N, Combescure C, Bouillon P. A speech-enabled fixed-phrase translator for emergency settings: crossover study. JMIR Med Inform 2019 May 07;7(2):e13167 [FREE Full text] [doi: 10.2196/13167] [Medline: 31066702]
27. Johnson M, Evans G, Mohamed Z, Caress A. The development and evaluation of alternative communication strategies to facilitate interactions with Somali refugees in primary care: a preliminary study. Inform Prim Care 2006 Sep 01;14(3):183-189 [FREE Full text] [doi: 10.14236/jhi.v14i3.629] [Medline: 17288704]
28. Villalobos O, Lynch S, DeBlieck C, Summers L. Utilization of a mobile app to assess psychiatric patients with limited English proficiency. Hisp J Behav Sci 2017 May 09;39(3):369-380. [doi: 10.1177/0739986317707490]
29. Arora S, Burner E, Lam J, De Santos R, Menchine M. Assessing the satisfaction of mobile health (mHealth) amongst ED inner-city patients with diabetes who received the TExT-MED intervention. Acad Emerg Med 2013;20(5(Suppl. 1)):S181.
30. Burner ER, Menchine MD, Kubicek K, Robles M, Arora S. Perceptions of successful cues to action and opportunities to augment behavioral triggers in diabetes self-management: qualitative analysis of a mobile intervention for low-income Latinos with diabetes. J Med Internet Res 2014 Jan 29;16(1):e25 [FREE Full text] [doi: 10.2196/jmir.2881] [Medline: 24476784]
31. Burner E, Menchine M, Taylor E, Arora S. Gender differences in diabetes self-management: a mixed-methods analysis of a mobile health intervention for inner-city Latino patients. J Diabetes Sci Technol 2013 Jan 01;7(1):111-118 [FREE Full text] [doi: 10.1177/193229681300700113] [Medline: 23439166]
32. Nápoles AM, Santoyo-Olsson J, Chacón L, Stewart AL, Dixit N, Ortiz C. Feasibility of a mobile phone app and telephone coaching survivorship care planning program among Spanish-speaking breast cancer survivors. JMIR Cancer 2019 Jul 09;5(2):e13543 [FREE Full text] [doi: 10.2196/13543] [Medline: 31290395]
33. Im EO, Ji X, Zhang J, Kim S, Lee Y, Chee E, et al. Issues in developing and evaluating a culturally tailored internet cancer support group. Comput Inform Nurs 2016;34(10):469. [doi: 10.1097/cin.0000000000000261]
34. Chee W, Lee Y, Im E, Chee E, Tsai H, Nishigaki M, et al. A culturally tailored Internet cancer support group for Asian American breast cancer survivors: a randomized controlled pilot intervention study. J Telemed Telecare 2017 Jul;23(6):618-626 [FREE Full text] [doi: 10.1177/1357633X16658369] [Medline: 27486198]
35. Kurth AE, Chhun N, Cleland CM, Crespo-Fierro M, Parés-Avila JA, Lizcano JA, et al. Linguistic and cultural adaptation of a computer-based counseling program (CARE+ Spanish) to support HIV treatment adherence and risk reduction for people living with HIV/AIDS: a randomized controlled trial. J Med Internet Res 2016 Jul 13;18(7):e195 [FREE Full text] [doi: 10.2196/jmir.5830] [Medline: 27417531]
36. Dahne J, Collado A, Lejuez C, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
37. Muroff J, Robinson W, Chassler D, López LM, Lundgren L, Guauque C, et al. An outcome study of the CASA-CHESS smartphone relapse prevention tool for Latinx Spanish-speakers with substance use disorders. Subst Use Misuse 2019 Mar 31;54(9):1438-1449. [doi: 10.1080/10826084.2019.1585457] [Medline: 30931681]
38. Muroff J, Robinson W, Chassler D, López LM, Gaitan E, Lundgren L, et al. Use of a smartphone recovery tool for Latinos with co-occurring alcohol and other drug disorders and mental disorders. J Dual Diagn 2017 Jul 10;13(4):280-290. [doi: 10.1080/15504263.2017.1348649]
39. Duracinsky M, Thonon F, Bun S, Ben Nasr I, Dara AF, Lakhdari S, et al. Good acceptability of HIV, HBV, and HCV screening during immigration medical check-up amongst migrants in France in the STRADA study. PLoS ONE 2020 Jun 26;15(6):e0235260. [doi: 10.1371/journal.pone.0235260]
40. Jancey J, Burns S, Hendriks J, Pollard CM, Tohatoa J, Hallett J. Measuring health promotion research impact - what researchers think? Health Promot J Austr 2020 Jun 26:1-7. [doi: 10.1002/hpja.379] [Medline: 32589313]
41. Torloni MR, Brizuela V, Betran AP. Mass media campaigns to reduce unnecessary caesarean sections: a systematic review. BMJ Glob Health 2020 Feb 26;5(2):e001935 [FREE Full text] [doi: 10.1136/bmjgh-2019-001935] [Medline: 32296554]
42. García-Álvarez D, Faubel R. Strategies and measurement tools in physical activity promotion interventions in the university setting: a systematic review. Int J Environ Res Public Health 2020 Sep 08;17(18):6526 [FREE Full text] [doi: 10.3390/ijerph17186526] [Medline: 32911659]
43. Charles Shapu R, Ismail S, Ahmad N, Lim PY, Abubakar Njodi I. Systematic review: effect of health education intervention on improving knowledge, attitudes and practices of adolescents on malnutrition. Nutrients 2020 Aug 13;12(8):2426 [FREE Full text] [doi: 10.3390/nu12082426] [Medline: 32823548]

XSL•FO
RenderX

44. Fagan MJ, Diaz JA, Reinert SE, Sciamanna CN, Fagan DM. Impact of interpretation method on clinic visit length. J Gen Intern Med 2003 Aug;18(8):634-638 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20701.x] [Medline: 12911645]
45. Tanser F, Bärnighausen T, Vandormael A, Dobra A. HIV treatment cascade in migrants and mobile populations. Curr Opin HIV AIDS 2015 Nov;10(6):430-438. [doi: 10.1097/COH.0000000000000192] [Medline: 26352396]
46. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealthtechnology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
47. Joseph RP, Royse KE, Benitez TJ. Systematic review of electronic and mobile health (e- and mHealth) physical activity interventions for African American and Hispanic women. J Phys Act Health 2019 Mar 01;16(3):230-239. [doi: 10.1123/jpah.2018-0103] [Medline: 30782040]
48. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-language consumer health information technology interventions: a systematic review. J Med Internet Res 2016 Aug 10;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
49. Pelikan JM, Krajic K, Dietscher C. The health promoting hospital (HPH): concept and development. Patient Educ Couns 2001 Dec;45(4):239-243. [doi: 10.1016/s0738-3991(01)00187-2]
50. Brender J, Talmon J, de Keizer N, Nykänen P, Rigby M, Ammenwerth E. STARE-HI - Statement on Reporting of Evaluation Studies in Health Informatics: explanation and elaboration. Appl Clin Inform 2013;4(3):331-358 [FREE Full text] [doi: 10.4338/ACI-2013-04-RA-0024] [Medline: 24155788]
51. Robustillo Cortés MDLA, Cantudo Cuenca MR, Morillo Verdugo R, Calvo Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. Telemed J E Health 2014 Aug;20(8):729-735. [doi: 10.1089/tmj.2013.0262] [Medline: 24849001]
52. Nykänen P, Kaipio J. Quality of health IT evaluations. Stud Health Technol Inform 2016;222:303. [doi: 10.3233/978-1-61499-635-4-291]
53. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth quality: a process to seal the qualified mobile health apps. Stud Health Technol Inform 2016;228:209.
54. Khander A, Farag S, Chen KT. Identification and evaluation of medical translator mobile applications using an adapted applications scoring system. Telemed J E Health 2018 Aug;24(8):594-603. [doi: 10.1089/tmj.2017.0150] [Medline: 29271702]
55. Mobile health app development costs $425,000 on average, likely continuing to rise. MobiHealthNews. 2018 Mar 26. URL: https://www.mobihealthnews.com/content/mobile-health-app-development-costs-425000-average-likely-continuing-rise [accessed 2021-04-12]

Abbreviations

CHEERS: Consolidated Health Economic Evaluation Reporting Standards
ICROMS: Integrated Quality Criteria for Review of Multiple Study Designs
HIV: human immunodeficiency virus
PRISMA: Preferred Reporting Items for Systematic Review and Meta-analysis

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by E Neter, K Schroeder; comments to author 02.12.20; revised version received 05.03.21; accepted 16.03.21; published 06.05.21.

Please cite as:
Thonon F, Perrot S, Yergolkar AV, Rousset-Torrente O, Griffith JW, Chassany O, Duracinsky M
Electronic Tools to Bridge the Language Gap in Health Care for People Who Have Migrated: Systematic Review
J Med Internet Res 2021;23(5):e25131
URL: https://www.jmir.org/2021/5/e25131
doi:10.2196/25131
PMID:33955837

©Frédérique Thonon, Swati Perrot, Abhijna Vithal Yergolkar, Olivia Rousset-Torrente, James W Griffith, Olivier Chassany, Martin Duracinsky. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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 Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure, Process, and Outcome

Rosian Tossaint-Schoenmakers[1,2,3], MSc; Anke Versluis[2,3], MSc, PhD; Niels Chavannes[2,3], MD, PhD; Esther Talboom-Kamp[2,3,4], MD, MBA, PhD; Marise Kasteleyn[2,3], MSc, PhD

[1]Saltro Diagnostic Centre, Utrecht, Netherlands

[2]National eHealth Living Lab, Leiden University Medical Centre, Leiden, Netherlands

[3]Public Health and Primary Care Department, Leiden University Medical Centre, Leiden, Netherlands

[4]Unilabs Group, Geneva, Switzerland

Corresponding Author:
Rosian Tossaint-Schoenmakers, MSc
Saltro Diagnostic Centre
Mississippidreef 83
Utrecht, 3565 CE
Netherlands
Phone: 31 6 55895361
Email: rtossaint@saltro.nl

Abstract

Background: Health care organizations are increasingly working with eHealth. However, the integration of eHealth into regular health care is challenging. It requires organizations to change the way they work and their structure and care processes to be adapted to ensure that eHealth supports the attainment of the desired outcomes.

Objective: The aims of this study are to investigate whether there are identifiable indicators in the structure, process, and outcome categories that are related to the successful integration of eHealth in regular health care, as well as to investigate which indicators of structure and process are related to outcome indicators.

Methods: A systematic literature review was conducted using the Donabedian Structure-Process-Outcome (SPO) framework to identify indicators that are related to the integration of eHealth into health care organizations. Data extraction sheets were designed to provide an overview of the study characteristics, eHealth characteristics, and indicators. The extracted indicators were organized into themes and subthemes of the structure, process, and outcome categories.

Results: Eleven studies were included, covering a variety of study designs, diseases, and eHealth tools. All studies identified structure, process, and outcome indicators that were potentially related to the integration of eHealth. The number of indicators found in the structure, process, and outcome categories was 175, 84, and 88, respectively. The themes with the most-noted indicators and their mutual interaction were inner setting (51 indicators, 16 interactions), care receiver (40 indicators, 11 interactions), and technology (38 indicators, 12 interactions)—all within the structure category; health care actions (38 indicators, 15 interactions) within the process category; and efficiency (30 indicators, 15 interactions) within the outcome category. In-depth examination identified four most-reported indicators, namely “deployment of human resources” (n=11), in the inner setting theme within the structure category; “ease of use” (n=16) and “technical issue” (n=10), both in the technology theme within the structure category; and “health logistics” (n=26), in the efficiency theme within the outcome category.

Conclusions: Three principles are important for the successful integration of eHealth into health care. First, the role of the care receiver needs to be incorporated into the organizational structure and daily care process. Second, the technology must be well attuned to the organizational structure and daily care process. Third, the deployment of human resources to the daily care processes needs to be aligned with the desired end results. Not adhering to these points could negatively affect the organization, daily process, or the end results.

(J Med Internet Res 2021;23(5):e27180) doi:10.2196/27180

XSL•FO
RenderX

KEYWORDS

eHealth; digital health; blended care; quality; integration; health care organization; structure; process; outcome

Introduction

Health care is changing, whereby patient empowerment, democratization of the internet, and an increasing burden on health care professionals play influential roles [1-3]. In line with these trends, innovations such as eHealth are required to maintain high quality of care [4-6]. eHealth includes a wide range of web-based interventions, for example e-consults, telemonitoring, and web-based viewing of medical records [1,7,8]. However, eHealth is more than a technology; it is another way of working and thinking and requires a change in attitude, which goes beyond the boundaries of a local health care organization [9,10].

The most comprehensive definition of eHealth with reference to the organizational context is that provided by Eysenbach [11]:

> *e-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology.*

In other words, the integration of eHealth into traditional health care requires organizational and behavioral changes for both health care professionals and patients [9,10].

Organizations are increasingly working with eHealth; however, implementing eHealth into the regular health care system requires organizations to change the way they work [9-11]. eHealth enables patients to have a more active role in managing their health [7,12,13], which affects interactions between the patient and health care professional [14-17]. Furthermore, working with eHealth technology requires workflow adjustments for health care professionals [18,19]. The organization's structure and care processes need to be adapted to ensure that eHealth supports the attainment of desired outcomes [20,21].

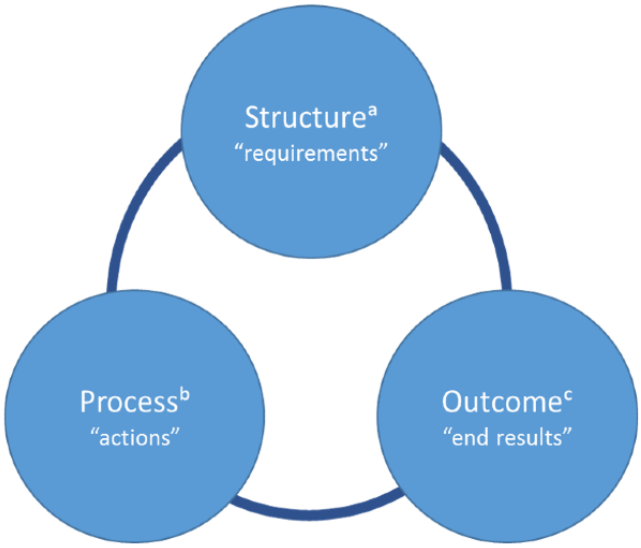
The challenge of optimally integrating eHealth into health care is thus a complex organizational issue. Several studies have identified elements to promote eHealth adoption, such as the degree of complexity, adaptability of the technology, costs, and stakeholder value [20,22], but uncertainty remains on how digital and traditional health care can blend successfully in the long term. With different definitions of eHealth available in the literature [10,11,23], and unclear barriers for facilitators in the application of eHealth [19], there is a need for further research on how eHealth can successfully be integrated into health care.

The aim of this study is to analyze how the integration of eHealth can be organized optimally by reviewing studies evaluating real-world eHealth interventions. The Donabedian framework of Structure-Process-Outcome (SPO) [24] was used, allowing the identification of relevant indicators that demonstrate how effective the integration of eHealth is in the organization.

According to the Donabedian model, the quality of health care can be assessed by three components that are relevant for organizations: structure (ie, requirements of the organization), process (ie, actions to be taken), and outcome (ie, end results), as shown in Figure 1 [24,25]. *Structure* is defined as the setting in which health care is provided (eg, facilities, equipment, numbers, and qualification of personnel); *process*, as what is actually done in giving and receiving care (eg, patient and doctor activities, doctor-patient communication and information); and *outcome*, as the consequence of the provided health care (eg, health status, satisfaction, and costs) [24-26]. Quality of health care is based on different aspects of these three categories and their relationships. As Donabedian eloquently puts it: "A good structure increases the likelihood of good process, and good process increases the likelihood of good outcomes" [24]. The interaction between the categories can be bidirectional, and it is not a simple separation between cause and effect [25]. The movement is an "unbroken chain of antecedents, followed by intermediate ends, which are themselves the means to still further ends" [25].

XSL•FO
RenderX

Figure 1. Donabedian Structure-Process-Outcome framework.



a. What an organization needs to have to provide health care
b. The actions in giving and receiving health care
c. End results as a consequence of providing care

The aim of this systematic review is twofold: (1) to investigate whether there are identifiable indicators in the structure, process, and outcome categories related to the successful integration of eHealth in regular health care and (2) to investigate which indicators of structure and process are related to outcome indicators.

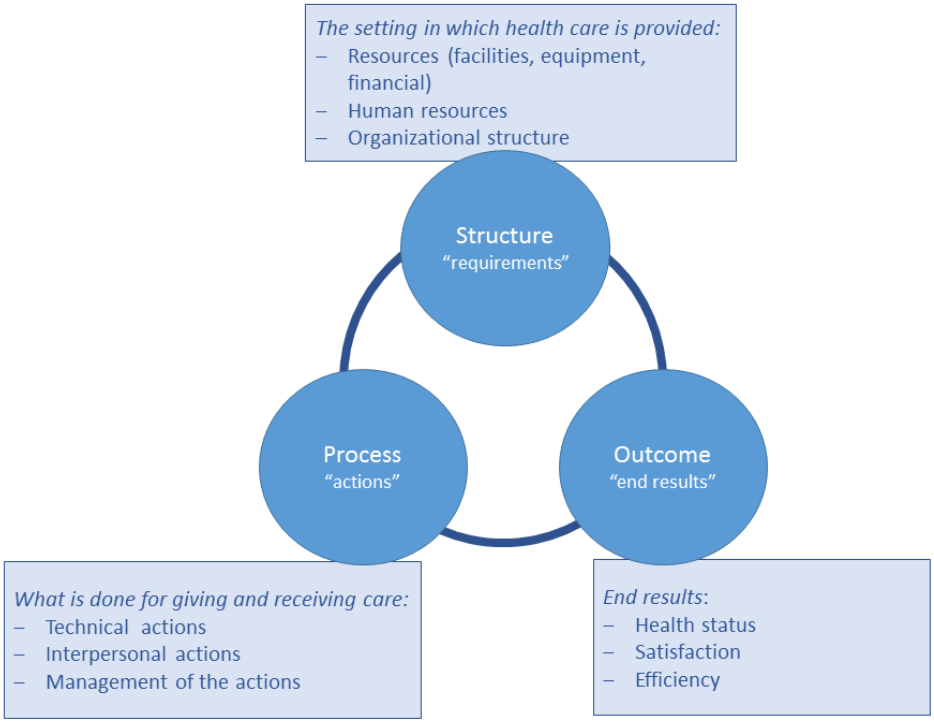
Methods

Theoretical Framework

The Donabedian SPO framework was used to identify the indicators of structure, process, and outcome that potentially affect the integration of eHealth into health care organizations. The Donabedian framework covers all relevant aspects of an organization's structure, process, and outcome and their interrelations, and it combines these aspects with health and social factors. Therefore, it is a suitable model to evaluate the organization of eHealth within health care organizations. The SPO categories are thematically explained in Figure 2 [24,25].

XSL•FO
RenderX

Figure 2. Explanation of the Structure-Process-Outcome categories of the Donabedian model.



The Donabedian SPO framework was designed in the 20th century before the introduction of eHealth. For this review, the SPO framework was adjusted to be compatible with the current time and incorporated the application of eHealth. The adjustments are described in the themes presented in Textbox 1.

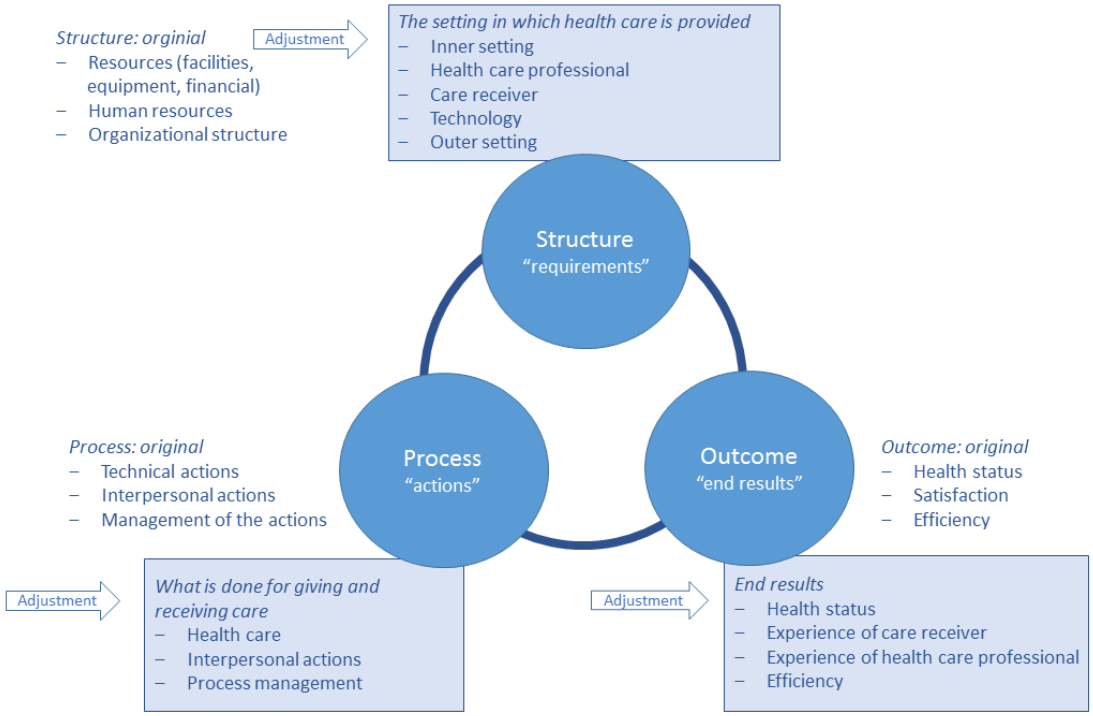
The adjustments to the SPO framework are shown in Figure 3.

Textbox 1. Adjustment of the Structure-Process-Outcome framework into themes, to integrate the application of eHealth into the health care system.

- **Structure**: The *setting* of provided care can be internal and/or external. Therefore, a distinction was made between *inner* and *outer settings*. With regard to *resources*, *technology* was added as a separate theme to cover eHealth. This was done because the focus of this research was eHealth. The remaining parts of the *resources* are covered under inner setting. *Human resources*, besides *health care professionals*, included *care receivers*. Their mutual involvement is required and is therefore also considered a conditional human resource [1]. *Organizational structure* was split into *inner setting* and *outer setting*, in line with the reasons given above, and to take the external stakeholders into account [27].
- **Process**: Instead of *technical actions*, the term *health care actions* was used, to avoid confusion with the term *technology* in the structure. *Interpersonal actions* remained unchanged. *Management of the actions* was shortened to *process management*.
- **Outcome**: *Health status* was retained as *health status*. *Satisfaction* was broadened to include *experience of the health care receiver* and *experience of the health care provider*, as both are pivotal outcome parameters in the *health care* process [28,29]. *Efficiency* remained unchanged.

XSL•FO
RenderX

Figure 3. Adjustment to the themes of the Structure-Process-Outcome framework, considering eHealth integration.



Search Strategy

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines. The research question was as follows: "How are structure indicators and/or process indicators related to eHealth or blended health care outcome indicators?"

Two authors (RT and MK) searched PubMed, EMBASE, Web of Science, Cochrane, and Emcare databases for relevant studies published up to December 12, 2019. They searched for the following terms in the titles, abstracts, and keywords of the published papers: structure* indicators* or process* indicators* or outcomes* indicators* and [blended care or eHealth* or telehealth*]. Multimedia Appendix 1 contains the full search details. After the search, two authors (RT and AV) screened the titles and abstracts of the relevant articles. Studies were included if they mentioned (1) the use of eHealth or blended care for diagnostics or treatment and (2) structure, process, or outcome indicators. Quantitative, qualitative, and mixed method study designs were included. A study was excluded if (1) it was a protocol, review, meta-analysis, grey literature, book chapter, oral presentation, or poster presentation; (2) it was published in a language other than English or Dutch; (3) full-text of the article was not available; (4) the intervention was not implemented (eg, conducted research regarding the users' expectations towards a prototype); or (5) the intervention used an analog application via plain-old-telephone lines. Of the remaining articles, RT reviewed the full texts. To ensure reliability, AV randomly selected about 10% of the fully reviewed articles for a blind review. Discrepancies were resolved by discussion. In case of uncertainty, a third author (MK) was consulted.

Data Extraction

Data extraction sheets were designed to provide an overview of the (1) study characteristics (eg, title, author, study aim, setting, disease, and quality appraisal); (2) characteristics of the eHealth intervention (eg, technology and function) and description of the intervention; (3) distribution of indicators into themes and categories related to the integration of eHealth into health care; and (4) interaction among the indicators, presented as themes.

RT designed the first concepts of the data extraction sheets. Authors RT, MK, NC, and ET discussed the design of the data extraction sheets to ensure their usability. Improved sheets were developed accordingly. The blind reviewer (AV) did not discuss the data extraction sheets. The included articles were reread by RT to check whether data clustering was complete and logical and for purposes of data pooling itself. AV selected a sample of 10% of the included articles for data extraction. Discrepancies were resolved by discussion.

Quality Appraisal

The Mixed Methods Appraisal Tool (MMAT) was used to appraise the quality of eligible studies in mixed methods systematic reviews—that is, reviews that included qualitative research, randomized controlled trials, nonrandomized studies, quantitative descriptive studies, and mixed methods studies [30]. The MMAT allows determination of the quality of different empirical study designs by using the same measure of five criteria in the chosen category. MMAT scores range on a scale of 1 to 5, with 1 indicating the lowest quality and 5 indicating the highest quality.

XSL•FO
RenderX

Classification of eHealth Interventions

eHealth interventions were ordered by type of technology and functionality. For technology, the classification proposed by Nictiz was used, distinguishing websites, apps, video communication, sensors, and wearables, domotics, robotics, and big data (ie, artificial intelligence) [31]. This classification is based on international studies [10,32]. For the present study, eHealth only concerns digital interventions and not analog ones such as analog applications via plain-old-telephone lines; this is in line with the classification proposed by Nictiz. For labelling the functionality, the second and third tiers of the National Institute for Health and Care Excellence (NICE) [33] were used, because these functionalities measure patient outcomes (Tier one consists of system services with no measurable patient outcomes). The functions were classified as communication, self-management, clinical calculation, active monitoring, diagnosis, and treatment [33].

Organization of Indicators and (Sub)themes of the SPO Framework

Indicators that had a potential impact on the integration of eHealth in health care were extracted and organized by the relevant theme according to the adjusted SPO framework (Textbox 1). In addition, the reported interactions between the indicators were extracted and organized by the relevant categories and themes. For a clear overview, the indicators within each theme were further divided into two subthemes by RT and ET (Table 1). The creation of subthemes was an iterative process. When reading the full texts, we found some definitions that sharpened some of the subthemes. The full definitions of the themes and subthemes are provided in Multimedia Appendix 2.

For each of the extracted indicators, the relevant impact on the integration of eHealth was noted. As there is no general standard for when eHealth is successful or effective [3,19], nor did the included studies specify such standards, these indicators were labeled as *advantage*, *disadvantage*, or *neutral*. An advantage in the structure and process categories indicates a positive effect on the integration and/or a positive effect on the outcome. A disadvantage in the structure and process categories indicates a negative effect on the integration and/or a negative effect on the outcome. An indicator that did not turn out to be an advantage nor a disadvantage was labeled neutral. The extracted indicators were noted as *advantage*, *disadvantage*, or *neutral*, in line with the evaluation performed in the corresponding study.

The following results are presented in this paper: (1) distribution of the indicators into (sub)themes and categories, and the impact on the integration of eHealth into health care; (2) most frequently reported indicators (ie, reported 10 times or more); and (3) interaction among indicators organized into themes and categories.

XSL•FO
RenderX

Table 1. Themes and subthemes in the structure, process, and outcome categories.

Category and theme	Subtheme
Structure	
Inner setting	• Support of primary process • Culture and leadership
Health care professional	• Skills • Attitude
Care receiver	• Daily life • Baseline characteristics
Technology	• Usability and functionality • Interaction with electronic health record
Outer setting	• Finance and legislation • Involvement of stakeholders
Process	
Health care actions	• Workflow • Patient-centered
Interpersonal actions	• Personal • Shifting roles
Process management	• Quality improvement • Mistake-proofing
Outcome	
Health status	• Clinical or functional • Intrapersonal
Experience of care recipient	• Satisfaction • Convenience
Experience of health care professional	• “What’s in it for me” • “What’s in it for them”
Efficiency	• Operations • Revenues

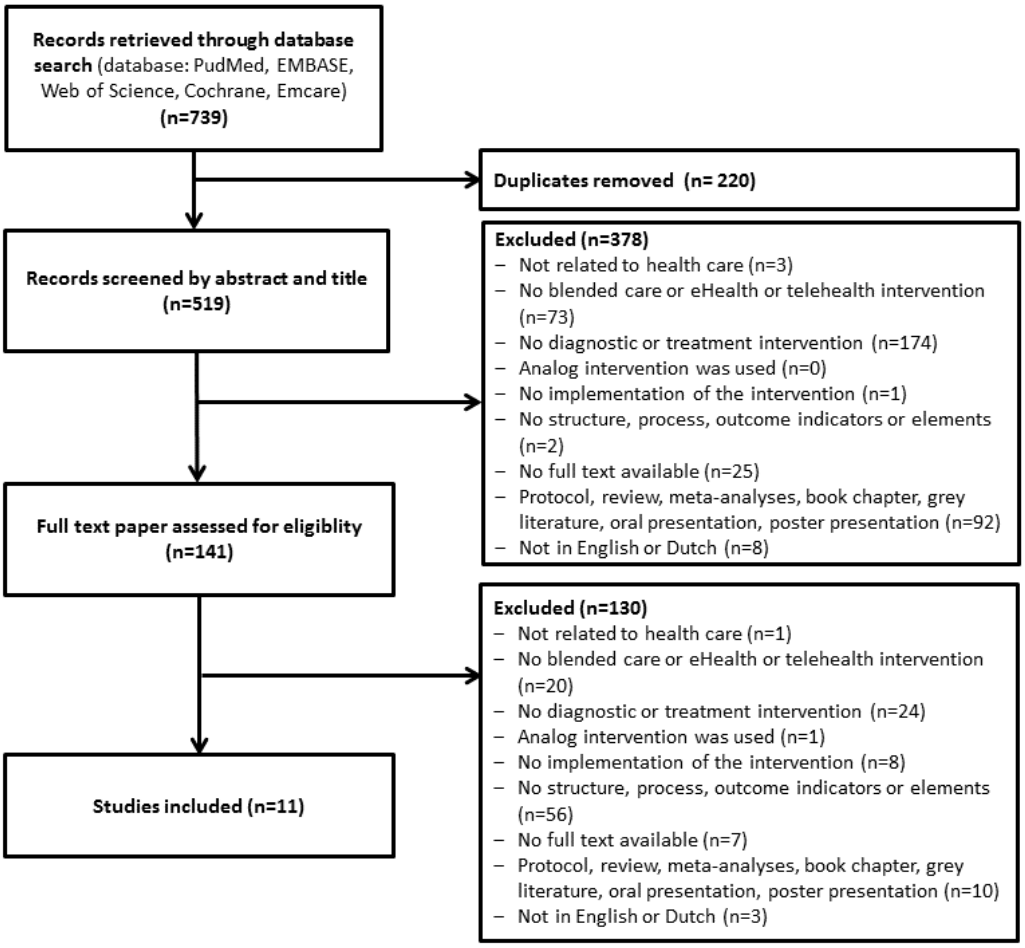
Results

Study Selection

The systematic search led to the identification of 11 eligible articles, selected from a total of 739 articles shortlisted initially (Figure 4).

XSL•FO
RenderX

Figure 4. Flowchart of the systematic review.



Data Results: Study and eHealth Characteristics

Study Characteristics

The included studies cover various study designs, diseases, and health care settings. Most studies were published after 2017 [27,34-41] and were of high quality [27,34,36,37,39,40,42]. Table 2 shows a detailed description of the characteristics of the included studies.

XSL•FO
RenderX

Table 2. Study characteristics.

Title, author (year)	Aim of the study	Study design	Setting and country	Disease	Sample size (N) and participant type	Quality of study[a]
Implementation of the blended care self-management program for caregivers of people with early-stage dementia, Boots et al (2017) [34]	To assess the internal and external validity of the trial and its implementation to inform effect analysis	Mixed methods, nested in an RCT[b]	Elderly care, home setting, the Netherlands	Early-stage dementia	N=62; Informal caregivers, psychologists, nurses	4.5
Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a Federally Qualified Health Centre, Thies et al (2017) [35]	To understand why the trial was unsuccessful	Qualitative (interview) analysis	Primary care, USA	Uncontrolled diabetes and hypertension	N=13; Patients, primary care provider, nurses, research assistants	3.0
"Sounds a bit crazy, but it was almost more personal:" A qualitative study of patient and clinician experiences of physical therapist-prescribed exercise for knee osteoarthritis via skype, Hinman et al (2017) [36]	To explore the experience of therapists and patients using Skype for exercise management of knee osteoarthritis	Qualitative study, nested in an RCT	Rehabilitation at home, Australia	Knee osteoarthritis	N=20; Patients, physical therapists	5.0
The challenge of real-world implementation of web-based share care software, Lycett et al (2014) [43]	To highlight the challenges of implementing software and reporting on the extent to which the software met its implementations and user aims	Mixed methods study, nested within an RCT	Children's Hospital, General Practices, Australia	Obesity	N=27; General Practitioners	2.5
Implementation of a multicomponent telemonitoring intervention to improve nutritional status of community-dwelling older adults, Van Doorn-van Atten et al (2018) [37]	To study how PhysioDom Home Dietary Intake Monitoring was delivered and received by participants and nurses, and to study the intervention's mechanism of impact	Mixed methods study	Home care and/or lived in a service flat in sheltered accommodation, the Netherlands	At risk of undernutrition	N=105; Patients, nurses	4.5
Implementation of internet-delivered cognitive behaviour therapy within community mental health clinics, Hadjistavropoulos et al (2017) [27]	To understand facilitators and barriers impacting the uptake and implementation of internet cognitive behavior therapy	Mixed methods study	Community Mental Health Clinic, Canada	Depression and anxiety	N=33; Therapists, managers	4.5
Implementation and evaluation of the safety net specialty care program in the Denver Metropolitan Area, Fort et al (2017) [38]	To describe the program, identify aspects that work well, areas for improvement, and offer lessons learned	Mixed methods study	Safety-net: a non-profit integrated health care system, USA	Uninsured patients	N=43; Patients, primary care clinicians, specialists	3.5
Perceived improvement in integrated management of childhood illness implementation through use of mobile technology, Mitchell et al (2012) [42]	To examine health care provider and carer perceptions of electronic Integrated Management of Childhood Illness (eIMCI) in diagnosing and treating childhood illnesses	Qualitative study (semi-structured interviews)	Health centers, Tanzania	Childhood illness in children 5 years or younger	N=31; Carers, health care providers	4.5

XSL•FO
RenderX

Title, author (year)	Aim of the study	Study design	Setting and country	Disease	Sample size (N) and participant type	Quality of study[a]
High level of integration in integrated disease management leads to higher usage in the e-Vita Study, Talboom-Kamp et al (2017) [39]	To analyze the factors that successfully promote the use of a self-management platform for chronic obstructive pulmonary disease patients	Quantitative, nonrandomized, parallel cohort design	Primary care, the Netherlands	Chronic obstructive pulmonary disease	N=215; Patients	4.0
What drives adoption of a computerised, multifaceted quality improvement intervention for cardiovascular disease management in primary healthcare settings? Patel et al (2018) [40]	To identify and explain the underlying mechanisms by which the intervention did and did not have an impact	Mixed methods study	Primary care, Australia	Cardiovascular disease	N=19; Patients, general practitioners, nurses, aboriginal health workers	4.5
Exploring the challenges of implementing a web-based telemonitoring strategy for teenagers with inflammatory bowel disease, Dijkstra et al (2019) [41]	To evaluate whether the telemonitoring strategy could move from a demonstration project to one that is sustained within existing sites	Mixed methods study, nested within an RCT	Pediatric gastroenterology centers, the Netherlands	Inflammatory bowel disease	N=27; Researchers, clinicians, hospital decision makers, web designer	2.5

[a]Methodological quality of studies assessed with MMAT, ranging from 0 (lowest) to 5 (highest).

[b]RCT: randomized controlled trial.

eHealth Intervention Characteristics, Descriptions, and Results

The most frequently used digital technology was websites (n=7) [27,34,37-41], and the most frequently reported functions [33] of the technology were self-management (n=6) [34-37,39,41] and communication (n=6) [35,37,39-41,43]. Table 3 shows an overview of the eHealth intervention characteristics, descriptions, and the study results. A detailed description of indicators, sorted according to the structure, process, outcome categories and their respective (sub)themes, are highlighted in the next paragraph.

XSL•FO
RenderX

Table 3. eHealth intervention characteristics, descriptions of the intervention, and study results.

First author	Technology; intervention name	eHealth function	Intervention	Study results (findings)[a]
Boots [34]	Website; Partner in Balance	Self-management	Face-to-face coaching with tailored web-based modules.	The participation rate of eligible caregivers was 51.9% (80/154). Recruitment barriers included lack of computer and need for support. Young age and employment were considered recruitment facilitators. All coaches attended training and supervision in blended care self-management. Deviations from the structured protocol were reported on intervention time, structure, and feedback. Coaches described an intensified relationship with the caregiver post-intervention. Caregivers appreciated the tailored content and positive feedback. The blended structure increased their openness. Overall, personal goals were attained after the program (t>50). Implementation barriers included lack of financing, time, and deviating target population.
Thies [35]	App; Undisclosed	Self-management, communication	Platform for active collaboration between patients and their primary care teams.	There was a poor fit between the app, end-users, and recruitment and treatment approaches in the setting. Usability testing might have revealed this prior to launch, but this was not an option. There was not sufficient time during routine care for clinical staff to familiarize patients with the app or to check clinical data and messages, which are unreimbursed activities. Some patients did not use the app appropriately. The lack of integration with the electronic health record was cited as a problem for both patients and staff who also said the app was just one more thing to attend to.
Hinman [36]	Video communication; Telerehabilitation via Skype	Self-management, treatment	Individualized home-based training strengthening program via Skype delivered physiotherapy.	Six themes arose from both patients and therapists. The themes were Structure: technology (ease of use, variable quality, set-up assistance helpful) and patient convenience (time-efficient, flexible, increased access); Process: empowerment to self-manage (facilitated by home environment and therapists focusing on effective treatment) and positive therapeutic relationships (personal undivided attention from therapists, supportive friendly interactions); and Outcomes: satisfaction with care (satisfying, enjoyable, patients would recommend, therapists felt Skype more useful as adjunct to usual practice) and patient benefits (reduced pain, improved function, improved confidence and self-efficacy). A seventh theme arose from therapists regarding process: adjusting routine treatment (need to modify habits, discomfort without hands-on, supported by research environment).
Lycett [43]	Website, HIE[b]; Shared-Care Obesity Trial in Children (HOPSCOTCH)	Communicate	Children attended a tertiary appointment with a pediatrician and dietician specializing in childhood obesity, followed-up by general practitioner consultations over the following year, supported by shared-care web-based software.	Software implementation posed difficult and at times disabling technological barriers. The software's speed and inability to seamlessly link with day-to-day software was a source of considerable frustration. Overall, general practitioners rated software usability as poor, although most (68%) felt that the structure and functionality of the software was useful.
Van Doorn-van Atten [37]	Website; PhysioDom HDIM[c]	Self-management, communication	Nutritional telemonitoring, education, a follow-up of telemonitoring measurement by a nurse.	About 80% of participants completed the intervention. Drop-outs were significantly older, had worse cognitive and physical functioning, and were more care-dependent. The intervention was largely implemented as intended and was received well by participants, but less well by nurses. Participants adhered better to weight telemonitoring than to telemonitoring by means of questionnaires, for which half the participants needed help. Intention to use the intervention was predicted by performance expectancy and social influence. No association was found between process indicators and intervention outcomes.

XSL•FO
RenderX

First author	Technology; intervention name	eHealth function	Intervention	Study results (findings)[a]
Hadjistavropoulos [27]	Website; ICBT[d]	Treatment	Web-based lessons that provide psychoeducation and instructions and therapist support via email or telephone.	ICBT implementation was perceived to be most prominently facilitated by intervention characteristics (namely, the relative advantages of ICBT compared to face-to-face therapy, the quality of the ICBT program that was delivered, and evidence supporting ICBT) and implementation processes (namely the use of an external facilitation unit that aided with engaging patients, therapists, and managers and ICBT implementation). The inner setting was identified as the most significant barrier to implementation as a result of limited resources for ICBT combined with greater priority given to face-to-face care.
Fort [38]	Website; Safety Net Specialty Care Program	Diagnosis, treatment	E-consults between primary care clinicians and specialist, face-to-face visits to the patients from a specialist, and continuing medical education for the primary care clinicians.	In the first 20 months of the program, safety-net clinicians at 23 clinics made 602 e-consults to specialists, and 81 patients received face-to-face specialist visits. Of 204 primary care clinicians, 103 made e-consults; 65 specialists participated in the program. Aspects facilitating program use were referral case managers' involvement and the use of clear, concise questions in e-consults. Key recommendations for process improvement were to promote an understanding of the different healthcare contexts, support provider-to-provider communication, facilitate hand-offs between settings, and clarify program scope.
Mitchell [42]	App; Electronic Integrated Management of Childhood Illness (eIMCI)	Diagnosis, treatment	An electronic handheld device or personal digital assistant, to guide the health care provider through IMCI protocols.	Providers expressed positive opinions on eIMCI, noting that the personal digital assistants were faster and easier to use than were the paper forms and encouraged adherence to IMCI procedures. Carers also held a positive view of eIMCI, noting improved service from providers, a more thorough examination of their child, and a perception that providers who used the personal digital assistants were more knowledgeable.
Talboom-Kamp [39]	Website; e-Vita COPD[e]	Self-management, communication	Insight into personal health data, self-monitoring of health values, education, and a coach for attaining personal goals.	Use of a self-management platform was higher when participants received adequate personal assistance about how to use the platform. Blended care, where digital health and usual care are integrated, will likely lead to increased use of the web-based program.
Patel [40]	HIE, website; HealthTracker	Communication, monitoring	Real-time decision support integrated with electronic medical records; CVD[f] risk communication tool between provider and patient; clinical audit tool; web portal providing peer-ranked performance trends.	A complex interaction was found between implementation processes and several contextual factors affecting uptake of the intervention. There was no clear association between team climate, job satisfaction, and intervention outcomes. There were four spheres of influence that appeared to enhance or detract from normalization of the intervention: organizational mission and history, leadership, team environment, and technical integrity of the intervention.

XSL•FO
RenderX

First author	Technology; intervention name	eHealth function	Intervention	Study results (findings)[a]
Dijkstra [41]	Website, IBD-live	Monitoring, self-management, communication	Flarometer, platform for direct communication with the IBD[g] team, module with study questionnaires (Quality of life, absenteeism, health care utilization).	The technology and the linked program allowed selection and targeting of teenagers who were most likely to benefit from a face-to-face encounter with their specialist. The value proposition of the technology was clear, with a distinct benefit for patients and an affordable service model, but health providers had plausible personal reasons to resist (double data entry). The organization was not yet ready for the innovation, as it required a shift to new ways of working. Dutch health insurers agreed that screen-to-screen consultations will be reimbursed at a rate equivalent to face-to-face consultations. The technology was considered easy to adapt and evolve over time to meet the needs of its users.

[a]Results [27,34-38,40-43] or conclusion [39] as described in the abstracts of the included studies.

[b]HIE: Health information exchange.

[c]HDIM: Home Dietary Intake Monitoring.

[d]ICBT: internet cognitive behavior therapy.

[e]COPD: chronic obstructive pulmonary disease.

[f]CVD: cardiovascular disease.

[g]IBD: inflammatory bowel disease.

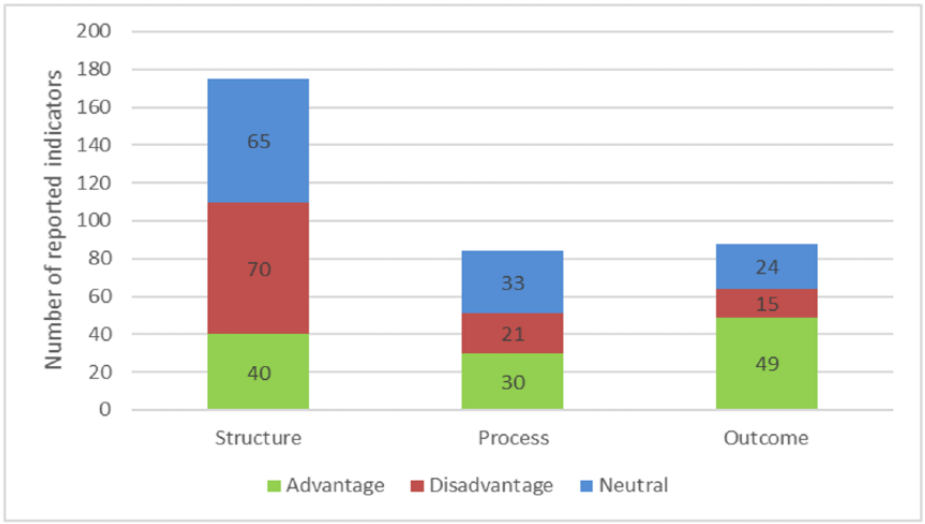
Indicators Organized by (Sub)themes of the SPO Framework

Overview

In total, an indicator was reported 347 times: 175 times in the structure category, 84 times in the process category, and 88 times in the outcome category. Of the 347 indicators, 111 were unique indicators (Multimedia Appendix 3). In the structure category, most indicators were labeled as neutral (65/175, 37.1%) or as a disadvantage (70/175, 40%). In the process category, most indicators were labeled as an advantage (30/84, 36%) or neutral (33/84, 39%). In the outcome category, the indicators were mostly classified as a realized advantage (49/88, 56%), as shown in Figure 5.

Table 4 shows the total distribution of the indicators organized by themes and subthemes of the structure, process, and outcome categories and the extent to which it was reported as an advantage, disadvantage, or neutral to the integration of eHealth and its outcome in regular health care. The themes and subthemes containing the most reported indicators are described below.

Figure 5. Number of indicators reported in the structure, process, and outcome categories. Advantage: in the structure and process categories, advantage indicates a positive effect on the integration. In the outcome category, it indicates a positive effect on the outcome. Disadvantage: in the structure and process categories, disadvantage indicates a negative effect on the integration. In the outcome category, it indicates a negative effect on the outcome. Neutral: Indicator was neither an advantage nor a disadvantage.



XSL•FO
RenderX

Table 4. Distribution of the indicators according to the themes and subthemes of the structure, process, and outcome categories.

Category, theme, and subtheme	Advantage (n)	Disadvantage (n)	Neutral (n)	Source
Structure (n=175)				
Inner setting (n=51)				
Support of primary process (n=34)	7	13	14	[27,34,37-43]
Culture and leadership (n=17)	7	9	1	[27,34,37,40]
Health care professional (n=28)				
Skills (n=8)	4	0	4	[27,36,38,40,41,43]
Attitude (n=20)	8	8	4	[27,34-41,43]
Care receiver (n=40)				
Daily life (n=18)	3	8	7	[27,34-39]
Baseline characteristics (n=22)	1	5	16	[34-39]
Technology (n=38)				
Usability and functionality (n=33)	8	17	8	[27,34-43]
Interaction with EHR[a] (n=5)	0	5	0	[35,37,38,41,43]
Outer setting (n=18)				
Finance and legislation (n=10)	0	2	8	[27,34,36,38-41]
Involvement of stakeholders (n=8)	2	3	3	[27,38,43]
Total structure	40	70	65	
Process (n=84)				
Health care actions (n=38)				
Workflow (n=18)	5	11	2	[27,34-39,41-43]
Patient-centered (n=20)	7	0	13	[27,34-39,41,42]
Interpersonal actions (n=24)				
Personal (n=19)	11	3	5	[27,34-42]
Shifting roles (n=5)	2	1	2	[34,36,42]
Process management (n=22)				
Quality improvement (n=11)	4	3	4	[27,34,38,40]
Mistake-proofing (n=11)	1	3	7	[27,37-39,41-43]
Total process	30	21	33	
Outcome (n=88)				
Health status (n=10)				
Clinical/functional (n=3)	1	0	2	[36,41,43]
Intrapersonal (n=7)	6	0	1	[34,36,37,41,42]
Experience of care receiver (n=23)				
Satisfaction (n=16)	11	3	2	[34-38,42]
Convenience (n=7)	7	0	0	[36,38,42]
Experience of health care professional (n=25)				
“What’s in it for me” (n=15)	9	2	4	[27,34,36-38,40,42]
“What’s in it for them” (n=10)	10	0	0	[27,34,36-38,41-43]
Efficiency (n=30)				
Operations (n=27)	4	9	14	[34-43]
Revenues (n=3)	1	1	1	[27,41,43]
Total outcome	49	15	24	

XSL•FO
RenderX

[a]EHR: electronic health record.

Distribution of Indicators Within the Themes and Subthemes of the Structure Category

In the structure category, most indicators were reported in the inner setting (51/175, 29.1%), care receiver (40/175, 22.9%), and technology (38/175, 21.7%) themes. The indicators in the inner setting (n=22) and technology (n=23) themes were mainly classified as a disadvantage to the integration, whereas those in the care receiver theme (n=23) were mainly classified as neutral. Regarding the subthemes, most indicators were reported in the support of the primary process subtheme within the inner setting theme (34/175, 19.4%), the baseline characteristics subtheme within the care receiver theme (22/175, 12.6%), and the usability and functionality subtheme within the technology theme (33/175, 18.9%), as shown in Table 4.

Distribution of Indicators Within the Themes and Subthemes of the Process Category

Almost half of the indicators were organized within the health care actions theme (38/84, 45%), which were diversely reported as an advantage (n=13), disadvantage (n=11), and neutral (n=15). The subthemes with the most reported indicators were workflow (18/84, 21%), patient-centered (20/84, 24%), both within the health care actions theme, and the personal subtheme (19/84, 23%) within the interpersonal actions theme (Table 4).

Distribution of Indicators Within the Themes and Subthemes of the Outcome Category

In the outcome category, the most frequently reported indicators were from the efficiency theme (30/88, 34%), with advantages (n=5) reported for very few indicators. The "experiences" themes of care receivers and health care professionals together accounted for 55% (48/88), both predominated by advantages (n=37). The highest number of indicators were reported in the operations subtheme (n=27/88, 31%; Table 4).

Most Reported Indicators

An in-depth examination of the distribution of the indicators showed that the following four indicators were the most reported (ie, reported 10 times or more) among the included studies: "deployment of human resources" (n=11) of the inner setting theme in the structure category; "ease of use" (n=16) and "technical issue" (n=10), both belonging to the technology theme in the structure category; and "health logistics" (n=26) of the efficiency theme in the outcome category. An overview of all indicators is presented in Multimedia Appendix 3.

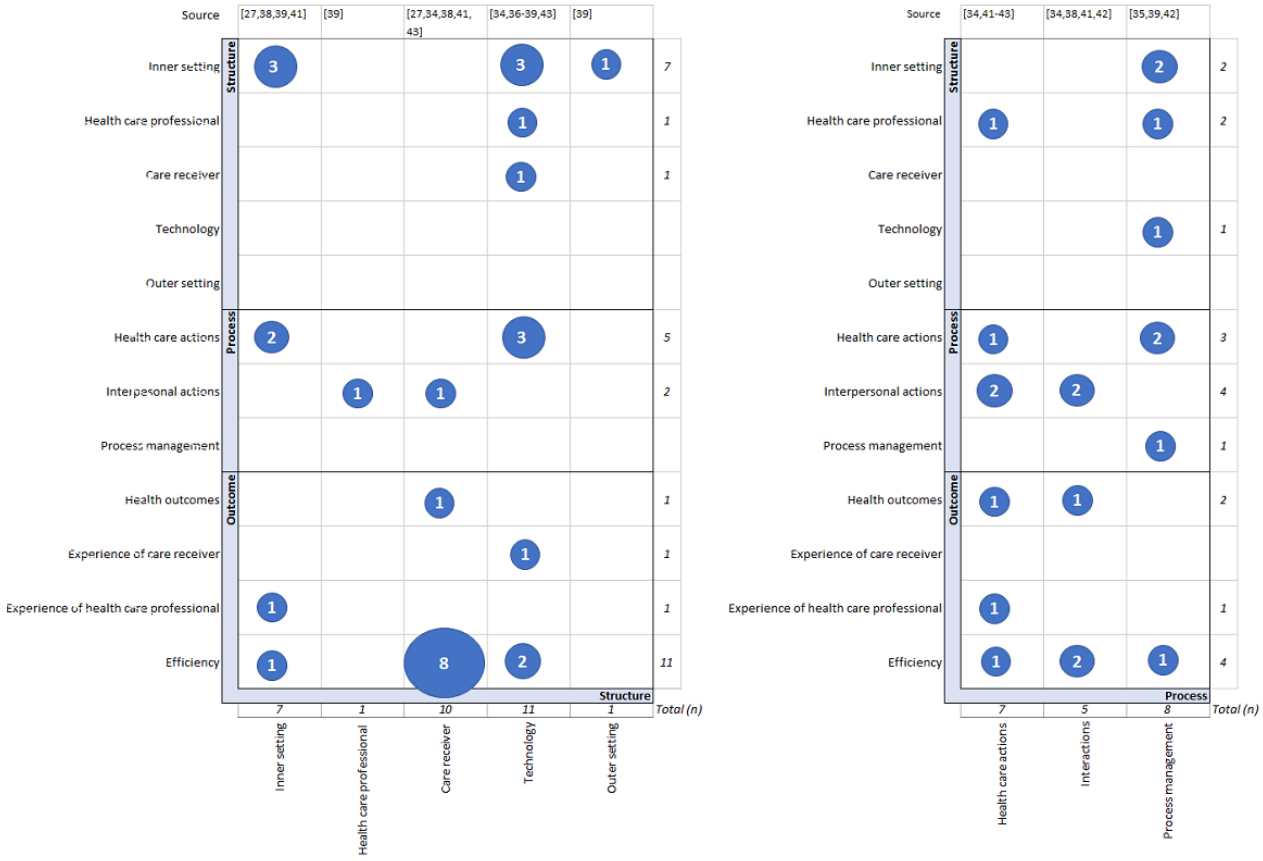
Interactions Among Indicators Organized into Themes and Categories

Overview

Of the 11 included studies, 10 (91%) reported interactions among indicators organized by themes within the structure, process, and outcome categories. The most frequently reported interaction among indicators at the category level was between the structure and outcome categories (14 times). The most frequently reported interaction among indicators at the theme level was between the care receiver theme within the structure category and the efficiency theme within the outcome category (8 times), as shown in Figure 6.

XSL•FO
RenderX

Figure 6. Interactions among indicators within themes and categories. The numbers within the blue circles represent the number of noted interactions among indicators within the themes. The x-axis represents the antecedent, and the y-axis represents the (intermediate) result.



Interactions With Themes in the Structure Category

All themes in the structure category contained indicators as an antecedent to, or as an intermediate result of other indicators. The inner setting (n=16), technology (n=12), and care receiver (n=11) themes represented the highest number of interactions with other themes. Inner setting was noted 7 times as an antecedent and 9 times as an intermediate result. Technology was noted 11 times as an antecedent and once as an intermediate result. Care receiver was noted 10 times as an antecedent and once as an intermediate result. The health care professional (n=3) and outer setting (n=1) themes were noted less frequently (Figure 6).

Interactions With Themes in the Process Category

In all themes in the process category, the indicators displayed interactions with indicators of other themes; specifically, health care theme (n=15), noted 7 times as an antecedent and 8 times as an intermediate result; interpersonal actions theme (n=11), 5 times as an antecedent and 6 times as an intermediate result; process management theme (n=9), 8 times as antecedent and once as an intermediate result (Figure 6).

Interactions With Themes in the Outcome Category

In the outcome category, the efficiency theme (n=15) contained most of the interacting indicators, all as an intermediate result. The other themes, including health status (n=3), experience of health care receiver (n=1), and experience of health care provider (n=2), were noted less frequently as (intermediate) results (Figure 6).

Examples of interactions among the indicators and the associated themes are illustrated in Textbox 2.

XSL•FO
RenderX

Textbox 2. Illustrations of reported interactions among indicators and their themes. Indicator names are written in italics as reported in the published studies (followed by the corresponding themes and categories in parentheses).

- *Technical and usability issues* (technology theme, structure category) experienced by the health care professional negatively impacted the *engagement* and the *internal collaboration* (inner setting theme, structure category) [40] and the health care workflow by causing *extra steps* and *workarounds* (health care actions theme, process category) [37,41,43].
- *Technical and usability issues* (technology theme, structure category) experienced by the care receiver challenged the care receiver to *fit the application of eHealth into their daily lives* (care receiver theme, structure category) and caused increased *dropouts* (efficiency theme, outcome category) [34,39]. Conversely, one study [36] showed that technology that is *easy to use* (technology theme, structure category), can contribute positively to its application, and *fit into the patient's daily life* (care receiver theme, structure category).
- Insufficient attention to the *patient's burden* (care receiver theme, structure category), *health literacy* (care receiver theme, structure category), and whether the plan *fits into their daily life* (care receiver theme, structure category) caused *dropouts* (efficiency theme, outcome category) [36,37,39], and *nonadherence to care plans* (efficiency theme, outcome category) [34].
- *High workload* (inner setting theme, structure category) hindered the *incorporation of the application into daily practice* (inner setting theme, structure category) [40].
- *Lack of time* (inner setting theme, structure category) discouraged health care professionals from their *intention to (re)*use (experience health care professional theme, outcome category) [37] and health care professionals did not *experience an added value for themselves* (experience health care professional theme, outcome category) [37].
- *Communicated added value* (inner setting theme, structure category) on a corporate level positively influenced the *collective engagement* (inner setting theme, structure category) [40].
- *Guidelines on the work process* (process management, process category) made the *work process easier and faster for health care professionals* (health care actions theme, process category) but *limited the adaptability of the technology for certain recipients* (technology theme, structure category) [42].
- Limited *feedback about the quality of care* (process management theme, process category) made specialists feel *uncertain about the suitability of the technology* (health care professional theme, structure category) [38], whereas *sharing information* (process management theme, process category) to improve program efficiency allowed the program *to be a part of the workflow* (health care actions theme, process category) [38].
- *Face-to-face contact* (health care actions theme, process category) benefitted the *personal connection between care receiver and professional* (interpersonal actions theme, process category) and the *engagement of the care receiver with the treatment* (interpersonal actions theme, process category) [34].
- *Personal assistance* (health care actions theme, process category) and *personalized therapy* (health care actions theme, process category) increased the *usage of the intervention by the care receiver* (efficiency theme, outcome category) [39].
- *Personalized therapy* (health care actions theme) also increased the *satisfaction of the care receiver* (experience of care receiver theme, outcome category) [36].
- *Exceptions to the operational process* (health care actions theme, process category) were made too often, such as *providing extra support to patients* (health care actions theme, process category), or *providing less care* (health care actions theme, process category), *creating new administrative workarounds* (health care actions theme, process category) caused by *technical issues* (technology theme, structure category) [35,37,38,41,43] or *high workloads* (inner setting theme, structure category) [27].
- An increase in *questioning* by professionals (interpersonal actions theme, process category) made carers feel more *engaged and knowledgeable* (health status theme, outcome category) [42].
- Recipients' *detailed input* (interpersonal actions theme, process category) on the assignments enabled professionals *to empathize with their situation* and *focus on their feedback* (interpersonal actions theme, process category) [34].

Discussion

Principal Findings

This literature review analyzed how eHealth can be organized optimally by using the Donabedian SPO framework. General organizational developments were identified, regardless of the type of illness, setting, or the eHealth application used. A review of the literature of selected cases highlighted three important findings. First, the role of the care recipient needs to be incorporated into the organizational structure and daily care process. Second, the technology must be well attuned to the structure of the organization and daily care process. Third, the deployment of the human resources to the daily processes needs to be aligned with the desired end results. Not adhering to these points could negatively affect the organization, daily process, or the end results. Findings from this research using the Donabedian framework corresponds to the conclusions of other studies using different research methodologies, which is explained below.

First, the SPO analysis showed that the care recipient plays a crucial role in the successful integration of eHealth. Patient-centered interaction and communication are important, to activate patients in managing their health care and to improve health outcomes in the application of eHealth [5,31,44-46]. Kuipers et al [44] and Rathert [45] demonstrated with systematic literature reviews that patient-centered care and co-creation are positively associated with the physical and social well-being of patients and with satisfaction of patients and health care professionals. These findings are in line with the review of Wildevuur and colleagues [46], demonstrating that organizations that are more patient centered with eHealth interventions achieve better outcomes with regard to patient health and quality of life.

XSL•FO
RenderX

Although most health care professionals embrace more patient involvement and engagement, delegating power and responsibilities could be a challenge for health care professionals' authority [47,48]. Another important issue is knowing who the customers are, what they want, and how the customer's demand is answered [49]. A previous study reported that eHealth is not suitable for all care receivers [18,50]. Therefore, identifying who benefits most from which kind of therapy is an essential addition to the screening process, and it could lead to more effective targeting and resourcing [51]. Furthermore, insufficiently incorporating the patients' family, work, and life goals into care plans will likely result in dropouts or nonadherence to care plans [50].

The second noteworthy finding is the essential role of excellent technology in the integration of eHealth. The way technology is set up has an influence on the working environment of health care professionals [52]. Inflexibility and complexity of the technology comes at the expense of effective daily processes and their quality [53,54]. Several studies demonstrated that the adaptability of eHealth technologies to fit to the local context, its ease of use, and its integration into clinical workflow benefit the users' acceptance and meaningful use [22,55,56]. This was also reflected in the early phase of the COVID-19 pandemic, where rapid scalable technologies were the easiest to use and quickly implementable [53]. However, the health care system continued to face challenges in adopting digital technology after the first emergency phase of the COVID-19 pandemic, due to inadequate information and communication technology infrastructure and a bad fit of the technology into the clinical workflow that is primarily designed for face-to-face care [53]. Granja et al [54] demonstrated that the application of eHealth is often not fitted to the existing workflow due to time and space constraints and breaking of traditions. Although eHealth is seen as an innovative solution for alleviating the increasing burden for health care professionals [2], it could have a counterproductive effect on the working conditions for employees if the technology is not properly adapted to the structure and processes [57,58]. Third, integrating eHealth into a health care organization requires adjustments of the care processes and utilization of the human resources, with appropriate process monitoring. Working with eHealth also poses logistical challenges; for example, a clear understanding is needed of the expected achievements, processes, and staffing requirements in order to bring about changes and create new capabilities [59]. Vissers and De Vries [49] pointed out that it is necessary to know how the logistical capacities should be assigned to the process, how the processes are measured, and who is responsible for the management of the process. Changes in the workflow are inevitable and necessary for eHealth interventions to be successful [54]. However, integrating eHealth technology into daily care processes is complex, and it needs coordination and process communication [19]. For example, a living laboratory experiment conducted over 3 years with patients, health care professionals, enterprises, and researchers to accelerate the integration of eHealth in daily practice showed that workflow, responsibilities, and roles needed to change, but health care professionals did not know how to approach this and had difficulties in integrating eHealth into their daily care processes [18].

Strength and Limitations

The strengths of this research are that international studies were included and represented a wide range of patient groups and settings. The findings were representative for the included studies, and they were not dependent of the study design, disease, target population, setting, or type or function of the eHealth application used. The wide range of settings of the included studies is supportive of a broader application of the present study's findings. In the *Methods*, we stated that there is no clear consensus on what constitutes as *good eHealth* and how it is best organized [3,19]. Nevertheless, we believe that our findings make a significant contribution to improving the integration of eHealth in regular health care by identifying the most common indicators in the organization's structure, processes, and outcomes. Thus, this research contributes to a new model for integrating organizational, health, and social factors.

A limitation of this study is that the health outcomes were rarely mentioned in this review. We hypothesized that this is because the main method used in the included studies was process evaluation. Therefore, although the health outcomes played a major role in earlier RCTs, this was not the case in process evaluation studies. The included studies did not define clear standards for the indicators to determine their quality. However, an indicator only becomes meaningful if a standard is specified [60,61]. There are also limitations in the selection procedure. The interrater reliability was not calculated. Due to this complex, broad topic, the predefined inclusion and exclusion criteria were sharpened at the time of selection. It was an iterative process, with a lot of consultation and coordination. In the process, full consensus was reached for all inclusion and exclusion criteria for selection at each step of the research. Another limitation is the classification of indicators into subthemes and themes at the discretion of the authors. It is conceivable that different classifications would reach different conclusions. Yet, the conclusions of each included study fit with the overall conclusion; therefore, the chance of this bias seems to be small. However, the findings of this literature review are dependent on the results of the included studies and may be subject to publication bias. Even though the included publications contain either positive or negative results (eg, a failed randomized trial [35] or interventions with no or less impact [40,43]), a chance of publication bias cannot be precluded automatically [62,63].

It is also noted that the Donabedian framework itself was designed before the introduction of eHealth and may not include the latest prevailing ideas on the organization of health care. For this reason, the model has been adapted in order to represent eHealth. By doing so, an attempt has been made to reduce the limitation as far as possible. Nevertheless, this literature review confirmed that it is still useful to analyze what contributes to the successful integration of eHealth into traditional health care. Additionally, there are other reputable models for evaluating eHealth interventions, such as the nonadoption, abandonment, scale-up, spread, sustainability (NASSS) framework [20]; Consolidated Framework for Implementation Research (CFIR) [64]; and the holistic framework to improve the uptake and impact of eHealth technologies [19]. These models describe the different phases from the design of the intervention to its

XSL•FO
RenderX

adoption and implementation. This literature review focused on quality improvement of the way eHealth is organized, that has already passed the initial phase (of design and adoption). The Donabedian framework covers all relevant aspects for sustaining the integration of eHealth into health care and the interrelations of organization's structure, processes, and outcomes, as well as integrating these aspects with human and social factors, after the adoption and uptake phase of eHealth.

Conclusions

For optimal integration of eHealth into health care, the following main principles should be considered and approached simultaneously. First, the role of the care recipient needs to be incorporated in the organizational structure and daily care process. Second, the technology must be well attuned to the structure of the organization and daily care process. Third, the deployment of human resources to the care process needs to be aligned with the desired end results.

Thus far, no study has presented a complete overview of the successful and effective organization of eHealth. Therefore, it is desirable to supplement this research with knowledge from other sources, such as in-depth research into the experiences from different perspectives, as this can help us to obtain a complete overview of how eHealth can be successfully integrated into health care organizations.

Conflicts of Interest

No conflict specified.

Multimedia Appendix 1
Search strategy.
[DOC File , 143 KB - jmir_v23i5e27180_app1.doc]

Multimedia Appendix 2
Explanatory notes on structure, process, and outcomes, and the (sub)themes.
[DOCX File , 20 KB - jmir_v23i5e27180_app2.docx]

Multimedia Appendix 3
Unique reported indicators.
[XLSX File (Microsoft Excel File), 22 KB - jmir_v23i5e27180_app3.xlsx]

References

1. van Lettow B, Wouters M, Sinnige J. E-health, what is that? Webpage in Dutch. Nictiz. URL: https://www.nictiz.nl/rapporten/e-health-wat-is-dat/ [accessed 2021-04-26]
2. Nijland N. Grounding eHealth: towards a holistic framework for sustainable eHealth technologies. University of Twente. 2011 Jan 21. URL: https://research.utwente.nl/en/publications/grounding-ehealth-towards-a-holistic-framework-for-sustainable-eh [accessed 2021-04-26]
3. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
4. Christensen CM. Disruptive innovation: can health care learn from other industries? A conversation with Clayton M. Christensen. Interview by Mark D. Smith. Health Aff (Millwood) 2007;26(3):w288-w295. [doi: 10.1377/hlthaff.26.3.w288] [Medline: 17355981]
5. Heckemann B, Wolf A, Ali L, Sonntag SM, Ekman I. Discovering untapped relationship potential with patients in telehealth: a qualitative interview study. BMJ Open 2016 Mar 02;6(3):e009750 [FREE Full text] [doi: 10.1136/bmjopen-2015-009750] [Medline: 26936904]
6. van der Kleij RMJJ, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJF, Teichert M, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
7. Talboom-Kamp E, Tossaint-Schoenmakers R, Goedhart A, Versluis A, Kasteleyn M. atients' attitudes toward an online patient portal for communicating laboratory test results: Real-world study using the eHealth impact questionnaire. JMIR Form Res 2020 Mar 04;4(3):e17060 [FREE Full text] [doi: 10.2196/17060] [Medline: 32024632]
8. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage. World Health Organization. Geneve, Switzerland: World Health Organization; 2016 Jan 01. URL: https://www.who.int/publications/i/item/9789241565219 [accessed 2021-04-29]
9. Chavennes N.H.. eHealth in Disease Management: doel of tool? 2015. https://scholarlypublications.universiteitleiden.nl/handle/1887/51560. Leiden, Netherlands: Leiden University; 2015 Dec 11. URL: https://hdl.handle.net/1887/51560 [accessed 2021-04-29]
10. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]

XSL•FO
RenderX

11. 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]
12. Matamala-Gomez M, Maisto M, Montana JI, Mavrodiev PA, Baglio F, Rossetto F, et al. The role of engagement in teleneurorehabilitation: A systematic review. Front Neurol 2020 May 6;11:354 [FREE Full text] [doi: 10.3389/fneur.2020.00354] [Medline: 32435227]
13. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. The Lancet Oncology 2020 May;21(5):e240-e251. [doi: 10.1016/s1470-2045(20)30021-8]
14. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015 Apr 24;10:59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
15. Boers SN, Jongsma KR, Lucivero F, Aardoom J, Büchner FL, de Vries M, et al. SERIES: eHealth in primary care. Part 2: Exploring the ethical implications of its application in primary care practice. Eur J Gen Pract 2020 Dec;26(1):26-32 [FREE Full text] [doi: 10.1080/13814788.2019.1678958] [Medline: 31663394]
16. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Numans ME, Chavannes NH. From chronic disease management to person-centered eHealth; a review on the necessity for blended care. Clinical eHealth 2018 Mar;1(1):3-7. [doi: 10.1016/j.ceh.2018.01.001]
17. Wildevuur SE, Simonse LW, Groenewegen P, Klink A. Information and communication technology enabling partnership in person-centred diabetes management: building a theoretical framework from an inductive case study in The Netherlands. BMJ Open 2019 Jun 16;9(6):e025930 [FREE Full text] [doi: 10.1136/bmjopen-2018-025930] [Medline: 31209085]
18. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
19. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
20. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
21. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
22. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
23. Boogerd EA, Arts T, Engelen LJ, van de Belt TH. "What is eHealth": Time for an update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
24. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]
25. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
26. Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? BMJ Qual Saf 2011 Apr;20(4):326-331. [doi: 10.1136/bmjqs.2010.042358] [Medline: 21339310]
27. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of internet-delivered cognitive behavior therapy within community mental health clinics: a process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]
28. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
29. Sikka R, Morath JM, Leape L. The Quadruple Aim: care, health, cost and meaning in work. BMJ Qual Saf 2015 Oct;24(10):608-610. [doi: 10.1136/bmjqs-2015-004160] [Medline: 26038586]
30. Hong Q, Pluye P, Fàbregues S. Mixed Methods Appraisal Tool (MMAT), Version 2018. User guide. Montréal, Canada;: McGill University; 2018 Aug 01. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf%0Ahttp://mixedmethodsappraisaltoolpublic.pbworks.com/ [accessed 2021-04-29]
31. Krijgsman GKWJ. Ordening in de wereld van eHealth (whitepaper). https://www.nictiz.nl/whitepapers/ordening-in-de-wereld-van-ehealth/. Den Haag, Netherlands: Nictiz; 2012 Aug 24. URL: https://www.nictiz.nl/whitepapers/ordening-in-de-wereld-van-ehealth/ [accessed 2021-04-29]

XSL•FO
RenderX

32. Meyer I, Hüsing T, Didero M, Korte W. eHealth Benchmarking (Phase II). Final report. Bonn, Germany; 2009 Mar. URL: https://joinup.ec.europa.eu/sites/default/files/document/2014-12/eHealth%20Benchmarking%20%28Phase%20II%29-Final%20Report.pdf [accessed 2021-04-29]
33. Cowie MR. National Institute for Health and Care Excellence. Eur Heart J 2015 Jan 21;36(4):195. [Medline: 25756143]
34. Boots LM, de Vugt ME, Smeets CM, Kempen GI, Verhey FR. Implementation of the blended care self-management program for caregivers of people with early-stage dementia (Partner in Balance): Process evaluation of a randomized controlled trialmplementation of the blended care self-management program for caregivers of people with early-stage dementia (Partner in Balance): Process evaluation of a randomized controlled trial. J Med Internet Res 2017 Dec 19;19(12):e423 [FREE Full text] [doi: 10.2196/jmir.7666] [Medline: 29258980]
35. 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]
36. Hinman RS, Nelligan RK, Bennell KL, Delany C. "Sounds a bit crazy, but it was almost more personal:" A qualitative study of patient and clinician experiences of physical therapist-prescribed exercise for knee osteoarthritis via Skype. Arthritis Care Res (Hoboken) 2017 Dec;69(12):1834-1844. [doi: 10.1002/acr.23218] [Medline: 28217864]
37. van Doorn-van Atten MN, de Groot LC, Romea AC, Schwartz S, de Vries JH, Haveman-Nies A. Implementation of a multicomponent telemonitoring intervention to improve nutritional status of community-dwelling older adults: a process evaluation. Public Health Nutr 2019 Feb;22(2):363-374. [doi: 10.1017/S1368980018002185] [Medline: 30175698]
38. Fort M, Namba LM, Dutcher S, Copeland T, Bermingham N, Fellenz C, et al. Implementation and evaluation of a safety net specialty care program in the Denver Metropolitan Area. Perm J 2017;21:16-022 [FREE Full text] [doi: 10.7812/TPP/16-022] [Medline: 28241908]
39. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Numans ME, et al. High level of integration in integrated disease management leads to higher usage in the e-Vita Study: Self-management of chronic obstructive pulmonary disease with web-based platforms in a parallel cohort design. J Med Internet Res 2017 May 31;19(5):e185 [FREE Full text] [doi: 10.2196/jmir.7037] [Medline: 28566268]
40. Patel B, Usherwood T, Harris M, Patel A, Panaretto K, Zwar N, et al. What drives adoption of a computerised, multifaceted quality improvement intervention for cardiovascular disease management in primary healthcare settings? A mixed methods analysis using normalisation process theory. Implement Sci 2018 Nov 12;13(1):140 [FREE Full text] [doi: 10.1186/s13012-018-0830-x] [Medline: 30419934]
41. Dijkstra A, Heida A, van Rheenen PF. Exploring the challenges of implementing a web-based telemonitoring strategy for teenagers with inflammatory bowel disease: Empirical case study. J Med Internet Res 2019 Mar 29;21(3):e11761 [FREE Full text] [doi: 10.2196/11761] [Medline: 30924785]
42. Mitchell M, Getchell M, Nkaka M, Msellemu D, Van Esch J, Hedt-Gauthier B. Perceived improvement in integrated management of childhood illness implementation through use of mobile technology: qualitative evidence from a pilot study in Tanzania. J Health Commun 2012;17 Suppl 1:118-127. [doi: 10.1080/10810730.2011.649105] [Medline: 22548605]
43. Lycett K, Wittert G, Gunn J, Hutton C, Clifford SA, Wake M. The challenges of real-world implementation of web-based shared care software: the HopSCOTCH Shared-Care Obesity Trial in Children. BMC Med Inform Decis Mak 2014 Jul 24;14:61 [FREE Full text] [doi: 10.1186/1472-6947-14-61] [Medline: 25056431]
44. Kuipers SJ, Cramm JM, Nieboer AP. The importance of patient-centered care and co-creation of care for satisfaction with care and physical and social well-being of patients with multi-morbidity in the primary care setting. BMC Health Serv Res 2019 Jan 08;19(1):13 [FREE Full text] [doi: 10.1186/s12913-018-3818-y] [Medline: 30621688]
45. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
46. Wildevuur SE, Simonse LWL. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
47. Tistad M, Lundell S, Wiklund M, Nyberg A, Holmner, Wadell K. Usefulness and relevance of an eHealth tool in supporting the self-management of chronic obstructive pulmonary disease: Explorative qualitative study of a cocreative process. JMIR Hum Factors 2018 Oct 26;5(4):e10801 [FREE Full text] [doi: 10.2196/10801] [Medline: 30368440]
48. Stoilkova-Hartmann A, Janssen DJA, Franssen FME, Spruit MA, Wouters EFM. Attitudes of healthcare professionals providing pulmonary rehabilitation toward partnership in care. Heart Lung 2015;44(4):347-352. [doi: 10.1016/j.hrtlng.2015.05.003] [Medline: 26025762]
49. Vissers J, Vries, de G. Sleutelen Aan Zorgprocessen. https://www.eur.nl/sites/corporate/files/Oratie_Jan_Vissers-Guus_de_Vries_Sleutelen_aan_Zorgprocessen_0.pdf. Rotterdam, Netherlands: Erasmus University; 2005 Apr 01. URL: https://www.eur.nl/sites/corporate/files/Oratie_Jan_Vissers-Guus_de_Vries_Sleutelen_aan_Zorgprocessen_0.pdf [accessed 2021-04-29]
50. Bellows J, Young S, Chase A. Person-focused care at Kaiser Permanente. Perm J 2014;18(1):90-91 [FREE Full text] [doi: 10.7812/TPP/13-165] [Medline: 24626077]

XSL•FO
RenderX

51. Newman S, Steed L, Mulligan K. Self-management interventions for chronic illness. Lancet 2004;364(9444):1523-1537. [doi: 10.1016/S0140-6736(04)17277-2] [Medline: 15500899]
52. van Merode F, van Raak A. Beheersing in de Zorg. Leerboek over Beheersingsvraagstukken Binnen Zorginstellingen En Zorgketens. In: Reed Business BV. Doetinchem, Netherlands: Reed Business BV; 2001.
53. 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]
54. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: Systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
55. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine Qualitative study. JMIR Res Protoc 2018 Dec 04;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
56. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. Perspect Health Inf Manag 2010 Oct 01;7:1d [FREE Full text] [Medline: 21063545]
57. Wouters M, Huygens M, Voogdt H, Meurs M, Groot, de J, Lamain A, et al. Samen aan zet! eHealth-monitor 2019. https://www.nivel.nl/nl/publicatie/samen-aan-zet-ehealth-monitor-2019. Den Haag and Utrecht, Netherland: Nictiz and het Nivel; 2019 Nov. URL: https://www.nivel.nl/nl/publicatie/samen-aan-zet-ehealth-monitor-2019 [accessed 2021-04-29]
58. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]
59. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
60. Zijlstra W. Quality Indicators in Healthcare. Webpage in Dutch. ZBC Kennisbank. Zeist, Netherlands; 2012 Jan 23. URL: https://zbc.nu/management/ontwikkeling-zorginstelling/kwaliteitsindicatoren-in-de-zorg/ [accessed 2021-04-29]
61. Harteloh P, Casparie A. Kwaliteit van zorg. Van Een Zorginhoudelijke Benadering Naar Een Bedrijfskundige Aanpak. In: P.M. H, Casparie AM. Maarssen, The Netherlands: Elsevier/ De Tijdstroom; 1998:144.
62. Müller KF, Briel M, D'Amario A, Kleijnen J, Marusic A, Wager E, et al. Defining publication bias: protocol for a systematic review of highly cited articles and proposal for a new framework. Syst Rev 2013 May 21;2:34 [FREE Full text] [doi: 10.1186/2046-4053-2-34] [Medline: 23692820]
63. Dickersin K, Min YI. Publication bias: the problem that won't go away. Ann N Y Acad Sci 1993 Dec 31;703:135-46; discussion 146. [doi: 10.1111/j.1749-6632.1993.tb26343.x] [Medline: 8192291]
64. 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 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
MMAT: Mixed Method Appraisal Tool
NASSS: nonadoption, abandonment, scale-up, spread, sustainability
NICE: National Institute for Health and Care Excellence
RCT: randomized controlled trial
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
SPO: Structure-Process-Outcome

Edited by G Eysenbach; submitted 14.01.21; peer-reviewed by C Basch, X Cheng; comments to author 27.01.21; revised version received 18.02.21; accepted 07.04.21; published 10.05.21.

Please cite as:
Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M
The Challenge of Integrating eHealth Into Health Care: Systematic Literature Review of the Donabedian Model of Structure, Process, and Outcome
J Med Internet Res 2021;23(5):e27180
URL: https://www.jmir.org/2021/5/e27180
doi:10.2196/27180
PMID:33970123

XSL•FO
RenderX

©Rosian Tossaint-Schoenmakers, Anke Versluis, Niels Chavannes, Esther Talboom-Kamp, Marise Kasteleyn. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.05.2021. 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 Use for Health Purposes: Systematic Review

Junhan Chen[1], MA; Yuan Wang[1], MA

Department of Communication, University of Maryland, College Park, MD, United States

Corresponding Author:
Junhan Chen, MA
Department of Communication
University of Maryland
Marie Mount Hall 0401
College Park, MD
United States
Phone: 1 608 572 9581
Email: jchen134@umd.edu

Abstract

Background: Social media has been widely used for health-related purposes, especially during the COVID-19 pandemic. Previous reviews have summarized social media uses for a specific health purpose such as health interventions, health campaigns, medical education, and disease outbreak surveillance. The most recent comprehensive review of social media uses for health purposes, however, was conducted in 2013. A systematic review that covers various health purposes is needed to reveal the new usages and research gaps that emerge in recent years.

Objective: This study aimed to provide a systematic review of social media uses for health purposes that have been identified in previous studies.

Methods: The researchers searched for peer-reviewed journal articles published between 2006 and 2020 in 12 databases covering medicine, public health, and social science. After coding the articles in terms of publication year, journal area, country, method, social media platform, and social media use for health purposes, the researchers provided a review of social media use for health purposes identified in these articles.

Results: This study summarized 10 social media uses for various health purposes by health institutions, health researchers and practitioners, and the public.

Conclusions: Social media can be used for various health purposes. Several new usages have emerged since 2013 including advancing health research and practice, social mobilization, and facilitating offline health-related services and events. Research gaps exist regarding advancing strategic use of social media based on audience segmentation, evaluating the impact of social media in health interventions, understanding the impact of health identity development, and addressing privacy concerns.

(J Med Internet Res 2021;23(5):e17917) doi:10.2196/17917

KEYWORDS

social media; health communication; health researchers; health practitioners; health institutions; systematic review

Introduction

Social media has been widely used in health contexts by various users. For individuals, social media is a primary source of COVID-19 information for patients [1]. Social media is used by 80% of cancer patients to connect with peers [2]. For health organizations, more than 80% of US state health departments have social media accounts [3]. Among health professionals, 65% of radiologists across the United States and Europe use social media for various health-related reasons [4]. A review of the state of the art can provide guidance for practitioners who intend to use social media and for researchers who intend to advance our understanding of social media use for health purposes.

Before reviewing social media use for health purposes, it is necessary to outline the scope of social media and health purposes. Social media has been defined in different ways. Some definitions focus on the technological features of social media that distinguish it from traditional technologies. For example, Kaplan and Haenlein [5] emphasized that social media is a type of application based on the internet and web 2.0 technology. Other definitions focus on the communication features of social media that distinguish it from traditional media. For example, McGowan et al [6] defined social media as an online

XSL•FO
RenderX

environment where users can contribute to the content and consume content mostly generated by other users. They also emphasized that content being "created by users for users" is an important feature that distinguishes social media from traditional media [6]. From a communication perspective, this paper emphasizes the communication features of social media and considers social media as a web 2.0–based platform for individuals to get access to, share, and generate content. Health purposes include health-related goals such as health promotion, medical service and administration, health research, medical education and training, and health-related social movements. This review focuses on users purposively getting access to, sharing, and generating content on social media in order to achieve various health purposes.

Previous reviews about social media use for health mostly focused on using social media for a single health purpose. These reviews summarized social media as a tool for health interventions [7], health campaigns [8], medical education [9], disease outbreak surveillance [10], and health promotion and behavior change [11]. The most recent comprehensive review of social media use for various health-related purposes was conducted by Moorhead et al [12] and included studies published between 2002 and 2012. However, multiple new uses have emerged between 2013 and 2020 given the rapid development of social media and the need for effective communication and infodemiology practice in the face of emerging health risks. Following Moorhead et al [12], this study reviewed articles published between 2006 and 2020 and categorized social media uses for health purposes in terms of user types, namely, health institutions, health researchers and professionals, and the public. By conducting a systematic review, we aimed to summarize social media use for health purposes, identify new uses that have emerged since 2013 as compared with Moorhead et al [12], and discuss research gaps and future directions for research in social media use for health purposes.

Methods

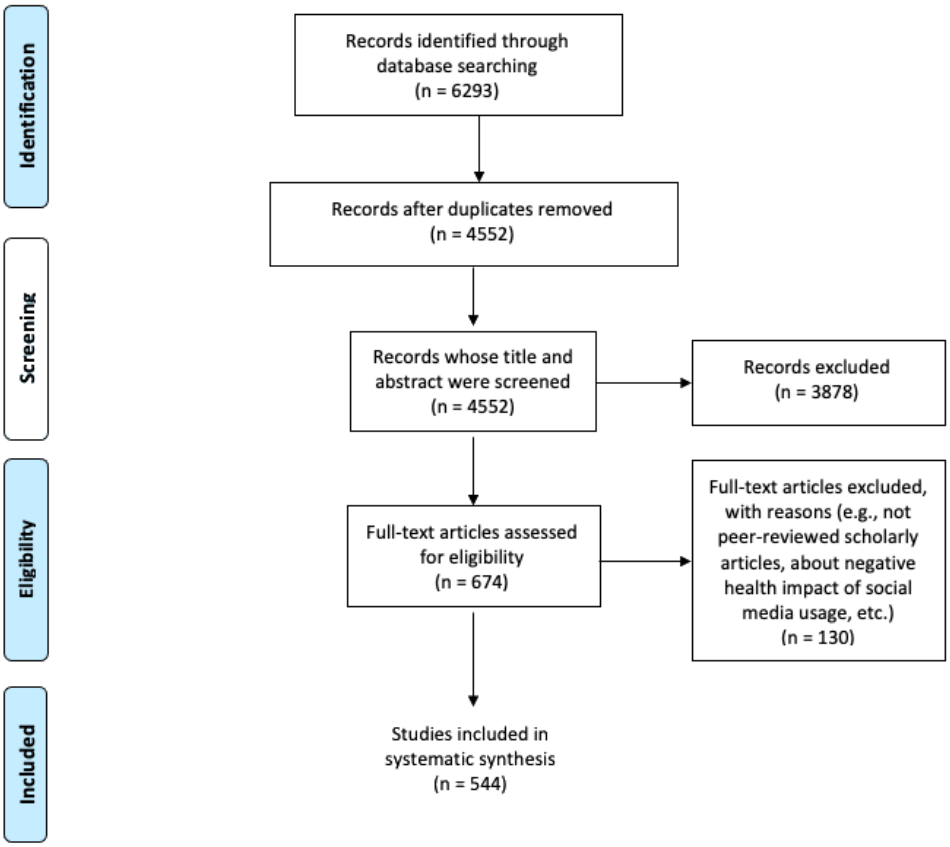
Searching Strategy and Inclusion Criteria

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guideline [13] to search and screen articles as shown in Figure 1. In the first step, the researchers searched 12 databases through ProQuest and EBSCO, including important databases in public health, medical, and social science areas such as MEDLINE, Academic Search Complete, PsycINFO, CINAHL, Psychology and Behavioral Science Collection, and Coronavirus Research Database. The search string was "AB(health) AND AB("social media" OR "social network" OR Facebook OR LinkedIn OR Twitter OR Instagram OR Pinterest OR WeChat OR Weibo)." Additional restrictions (ie, published between 2006 and 2020, peer-reviewed scholarly journal article, full text available, and written in English) were added to the search requirement. The time range was set from 2006 to the end of 2020 because most of the current popular social media platforms such as Twitter and Facebook have been created or open to the public since 2006. The initial search retrieved 6293 articles.

In the next step, the two researchers screened items retrieved from the databases. An article was included if the study identifies at least one social media use for health purposes (eg, a study that finds that parents use social media to seek information about breastfeeding), examines the impact of using social media for health purposes (eg, a study that examines the effectiveness of a social media–based health intervention), or explores how to use social media for health purposes (eg, a study that examines how to detect mental health risk through analyzing social media posts). After reviewing the titles, abstracts, and full text, the researchers included 544 articles in the final sample. Articles were excluded for reasons such as not being scholarly journal articles, focusing on the negative health impacts of using social media such as depression and obesity, describing social media networks or social media post characteristics without discussing how to intentionally use social media for health purposes, examining health promotion within an offline social network, or examining social media use in general context rather than health context.

XSL•FO
RenderX

Figure 1. Article search and screening process.



Coding Process

The preliminary coding scheme was developed by one researcher based on an initial review of 20% of the articles included. The two researchers coded 20 articles using the preliminary coding scheme and met to discuss the applicability of the preliminary coding categories and possible additional coding categories. The coding scheme was finalized after the two researchers reached an agreement after the discussion. The finalized coding scheme included 8 dimensions: publication year, journal area (eg, public health and medical area, communication, psychology), the country where the study was conducted (eg, United States, United Kingdom, China), type of article (eg, empirical studies, review paper, protocol), methodology used in empirical studies (eg, experiment, survey, interview), type of social media users (ie, health institution, health researcher and professional, publics), and social media use for health purposes. Social media use for health purposes were grouped under social media users. There are 4 usages by health institutions (ie, infoveillance, disseminate health information and combat misinformation, health intervention, and social mobilization), 3 usages by health researchers and professionals (ie, health-related research, professional development, and facilitate doctor-patient communication and offline services), and 3 usages by the public (ie, seek and share health information, exchange social support in an online community, and track and share one's health statuses or activities). The intercoder reliability indicated by Krippendorff alpha ranged from a low of 0.86 to a high of 1.00. The reliability is satisfactory compared with the 0.80 criterion [14].

Results

Characteristics of the Studies

All articles were published between 2016 and 2020. There were 25 articles published in 2016, 42 in 2017, 28 in 2018, 185 in 2019, and 264 in 2020. The number of publications continuously increased during the 5 years, showing that social media use for health purposes received growing attention from researchers.

Regarding methodology, most of the articles were empirical studies (n=399). Among the empirical studies, 294 used a quantitative method including experiment, survey, content analysis, and network analysis, and 74 studies used a qualitative method including interview, focus group, and discourse/thematic analysis. There were 15 studies using a mixed method and 16 studies used other methods such as data mining and describing social media metrics of campaign messages (eg, number of likes and views). The rest of the articles were review papers (n=78), commentaries (n=37), protocols (n=22), and theoretical and methodological papers (n=8).

Regarding fields of publication, most articles were published in journals in medical and public health (n=489) followed by journals in communication (n=26), psychology (n=7), sociology (n=6), information science (n=5), interdisciplinary journals (n=5), and journals of other fields (n=6) such as technology, computing, and behavior research method.

Most of the articles featured social media use in the United States (n=170) followed by China (n=64), United Kingdom (n=20), Australia (n=14), Canada (n=7), South Africa (n=6),

XSL•FO
RenderX

Iran (n=5), Malaysia (n=4), Ghana (n=4), and India (n=4). A total of 43 articles featured social media use in other countries such as Jordan, South Korea, Denmark, and Sweden. Some studies featured social media use in more than one country or internationally (n=27). There were 176 studies that did not specify the geographic scope of social media use. For example, there were studies that analyzed all the English written posts related to a disease on social media or reviewed all the relevant literature about social media–based health intervention.

Regarding the social media platforms used for health purposes, most studies (n=231) examined social media in general without specifying use on certain social media platforms. Other studies featured Twitter (n=79) followed by Facebook (n=76), WeChat (n=33), online forums (n=16), Sina Weibo (n=13), Reddit (n=8), YouTube (n=8), WhatsApp (n=6), Instagram (n=6), and other platforms such as Pinterest, Yelp, and Yahoo! Answer (n=10). There were also 58 studies that featured more than one social media platform.

Social Media Use for Health Purposes

This review summarized 10 uses of social media for health purposes. Uses were grouped into 3 categories according to the type of social media users, namely, health institution, health researcher and professional, and the public. The following sections summarize social media uses by each type of user and the benefits and challenges related to each use.

Health Institution

Health institutions include government health agencies and nongovernmental health organizations such as the World Health Organization. The review shows that health institutions mainly used social media for infoveillance (n=94), disseminating health information and combating misinformation (n=76), health intervention (n=168), and social mobilization (n=7).

Infoveillance

Infoveillance is the application of infodemiology with a primary aim of surveillance, which refers to surveilling and analyzing the unstructured information available on the internet in order to inform public health and public policy [15]. In this review, we focused on the surveillance of information available on social media in order to inform public health. Specifically, articles in this category analyzed social media data to (1) surveil potential public health risks including the onset of individual illnesses and disease outbreaks in a population (n=70) and (2) monitor public responses to health issues (n=25).

First, social media data have been analyzed to predict the future illness onset for individual users. For example, the language used in users' posts such as life stress and anxiety can predict future occurrence of mental illness and distinguish different types of mental illness [16]. In addition, special characters used in social media posts such as emojis can also help identify illness syndromes such as asthma and difficulty breathing syndrome [17]. The use of social media to detect individual illness is especially beneficial for people who are at risk but less motivated to seek diagnosis and treatment; social media surveillance can supplement the traditional doctor-patient interaction method for illness detection [17] and can be combined with social media–based interventions to proactively provide health advice to those who are less likely to seek help [18].

Second, social media data can be analyzed to predict infectious disease outbreaks among a population such as in the COVID-19 outbreak [19]. The studies showed that social media data can provide an accurate prediction of disease outbreak case count [19]. In addition to outbreak prediction, demographic and geographic data obtained from social media can inform medical research and practice of the characteristics of people who are at risk of being infected [20]. Consistent with a recent systematic review of social media–based surveillance systems for health care [21], this study found that Twitter was the most used social media platform for individual illness and disease outbreak surveillance.

Third, user-generated social media posts can provide insights about the public's cognitive and behavioral responses to health-related issues. Studies have analyzed social media posts to understand public concerns and sentiments about ongoing disease outbreaks such as Zika [22] and COVID-19 [23], investigated public attitudes toward and actual implementation of recommended disease prevention actions [20,24], monitored public discussion about controversial health topics such as vaccination [25] and e-cigarettes [26], and gauged public support of health-related policies [27]. Assessing social media data to understand public responses to health issues not only yields results comparable to national survey data but also provides insights for health-related theoretical models [25]. More importantly, traditional public surveys may take weeks and are expensive, while social media provides real-time and inexpensive data to evaluate the effectiveness of public health communication, which allows communication practitioners to adapt timely communication strategies to public needs [23].

Disseminate Health Information and Combat Misinformation

With the ability to disseminate information quickly and widely, social media can be used by health institutions to post health information and share information with the public. Health institutions can inform the public about common health topics such as healthy living, immunization, and smoking [7]; communicate the risk of disease outbreaks [28]; quickly provide instructions about prevention behaviors to a broad audience during disease outbreaks [29]; and share the latest news and inform the public of the government's handling of disease outbreaks [28,30]. Research efforts have been devoted to exploring strategies for generating health content on social media and increasing retransmission [31,32].

Social media has several advantages over other media outlets when used for disseminating health information. First, social media is seen as the fastest among available channels to share alerts and updates about disease outbreaks [28]. Second, social media allows leveraging various media forms to engage the public. For example, by integrating hyperlinks, social media posts can direct the public to other online resources for additional health information [28,33]. Health institutions can also share podcast audios and YouTube videos on various social media platforms to deliver health information [34,35]. Third, social media can provide a communication platform for

XSL•FO
RenderX

stakeholders during disease outbreaks. Government social media accounts may become official information sources that provide content about disease outbreaks for local agencies and journalists in a timely manner [28,33].

With the increased volume of misinformation during the COVID-19 pandemic, many studies focused on the concern of rapid misinformation transmission on social media and proposed approaches to combat misinformation through social media. For example, researchers suggested health institutions increase efforts to monitor and enforce fact-checking of dubious information on social media [36,37], promote strategic and timely refutation to misinformation on social media [38], share personal experiences on social media to refute rumors [39], and use search optimization strategies within the social media platforms to redirect users who ask health-related questions to reliable information sources [40]. In addition to combating misinformation through social media, health institutions may also encourage health professionals to establish interpersonal relationships with patients and invite patients to conversations about misinformation that they may encounter online [41]. Although the effectiveness of the approaches proposed by articles in this review have not been formally tested, there have been many studies testing the effectiveness of different message design features in correcting misinformation. For example, Gesser-Edelsburg and colleagues [42] found that a theory-based correction message posted on Facebook by health authorities was more effective in correcting vaccination misinformation than a common correction message. Practitioners may design and post misinformation-correcting messages on social media based on the strategies suggested by these message design studies.

Health Intervention

Social media has been widely integrated into health interventions and can perform multiple functions. First, social media was used to offer health resources including delivering health information to audiences [43], motivating participation in health-related events [44], and directing campaign audiences and intervention participants to other health resources [45,46]. Second, social media was used to support interaction with audience including engaging campaign audiences in 2-way communication with the institutions and health professionals [47,48] and providing platforms for peer support group discussion [49]. Third, social media was used to motivate health behavior change; intervention participants used social media to document and share their own progress of health behavior change such as uploading a picture of an everyday healthy meal [50] and sharing virtual awards of weekly achievement in physical activities [51], setting group challenges toward health goals [52], and engaging in health behavior competition with peers [53]. Last, at the preparing stage of an intervention, social media was also used to reach broader and more varied populations, including hard-to-reach populations [54,55].

Research effort has also focused on improving the design and evaluation of social media–based health interventions. Studies have examined approaches to evaluating campaign outcomes [56], proposed theoretical frameworks for effective health interventions design [57], explored social media user segments to inform intervention development [58], and examined user acceptance and suggestions of receiving social media–based interventions [59]. In general, empirical evidence shows that integrating social media in health interventions is accepted by target populations [59] and effective in increasing health knowledge, reducing risky behaviors, and adopting health behaviors [60]. Researchers have also discussed challenges and raises concerns regarding social media–based health intervention; for example, researchers must rigorously evaluate social media–based interventions throughout the stages of development and implementation, involve populations experiencing health disparities in the studies, and address privacy concerns [61].

Social Mobilization

Since social media can reach varied stakeholders and a wide audience in society, nongovernmental health organizations can use social media to mobilize social resources. Studies found that these organizations leveraged social media to advocate for change in public policies related to health issues such as HIV/AIDS [62] and mental health [63], raise funds for individual medical care and health-related research [64,65], and raise awareness and promote actions to address health problems such as medical equipment shortages [66]. Although a formal test of the effectiveness of using social media for social mobilization is lacking, the successful cases described in the articles show that social media provides a platform for nongovernmental health organizations to effectively mobilize resources and advocate for collective actions to achieve their goals [62].

Health Researchers and Practitioners

Social media use by health researchers includes facilitating health-related research (n=45), professional development (n=34), and doctor-patient communication and offline services (n=36).

Facilitate Health-Related Research

Social media can facilitate health-related research in two ways: provide additional data to learn about patients' disease experience and recruit research participants. First, by analyzing patients' conversations on social media, researchers can gauge patients' understanding of the disease and their coping strategies [67], identify their concerns about the disease [68], understand their barriers to health behavior change [69], identify symptoms related to the disease [70], and assess patients' experience after recovery [71]. Because patient self-reported illness experience is not routinely reported to and recorded by physicians, analyzing patient discussions on social media may enhance health researchers' and professionals' understanding of patient experiences beyond what can be learned from traditional sources of health data [72].

Second, social media can help researchers recruit participants for health-related research. Studies found that social media performs as well as or even better than traditional recruiting methods in terms of cost and number of participants enrolled [73]. Social media is also useful in recruiting hard-to-reach populations such as immigrants [74]. Social media recruitment can generate a representative sample; a study comparing a social media–recruited sample with the underlying population found

XSL•FO
RenderX

that the sample was representative in 8 of the 13 characteristics studied including demographics and health-related factors [75].

Professional Development

Health professionals and researchers may use social media for their own professional development such as learning, collaboration, and career advancement. Social media can be used to collaborate on research projects and practices [76], access and share trending research findings and medical knowledge [77,78], broaden their exposure to funders and publishers [79], conduct a job search [80], follow medical conferences remotely [81], market their team and services, and discuss interesting or difficult cases with colleagues [4]. Studies found that different social media platforms serve different roles in the professional development of health professionals and researchers; Pinterest was mostly used for health care–quality education [82], Twitter for gathering news and information on conferences, and LinkedIn for career advancement [81]. Nearly 85% of health professionals agreed that social media can be an effective tool for educational purposes [82], and 71% of health professionals, researchers, and businesspeople in the urology discipline agreed that social media is useful for career development [83].

Facilitate Doctor-Patient Communication and Offline Health Services

Health professionals use social media for doctor-patient communication including responding to questions posted by patients [84], offering online consultation [85], and proactively providing advice and health information to social media followers [86]. In addition, social media can be used to support offline health services. Health professionals can use social media to inform patients of the results of their examination [4], encourage compliance with medication [87], receive feedback from patients about their health services [4], and collect information from patients after discharge to inform future practices [88]. Encouraging interactions between health practitioners and patients on social media may benefit both patients' well-being and patient-physician relationships [89]. In addition, social media can also be used to provide medical services such as appointments, medical inquiries, personal information management, and medical charge payment, which can increase the accessibility of medical care and improve patient experiences [90].

However, increasing use of social media by health professionals and researchers is accompanied by concerns and challenges. Many studies raised concerns about risks to patients' privacy and confidentiality [80] and health professionals lacking knowledge about social media [4]. Studies also discussed challenges such as protecting content generators' intellectual property [91] and managing negative comments from patients [4]. Most of these articles called for developing a guideline or a regulation regarding ethical, legal, and technological issues [4,80] and providing training for health professionals and researchers on the proper use of social media [91,92].

Public

The public includes healthy people and those with health concerns, including patients with known diagnoses and populations at risk of certain health problems, such as men who have sex with men, first-time pregnant mothers, and LGBTQIA+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual) youth. The reviews show that the public uses social media to seek and share health-related information (n=108), exchange social support in online communities (n=95), and track and share health statuses or activities (n=10).

Seek and Share Health-Related Information

Seeking and sharing health information is the most common use by the public. The public uses social media to follow and share news about trending health issues such as flu [93] and COVID-19 [1], find information on daily health behaviors such as fitness and healthy food choices [94-96], search for health advice for their own health concern such as pregnancy-related information [97], request a crowd diagnosis or second opinion after seeing a health care professional about their illness [98], access health care law [70], follow social media accounts of health organizations and professionals [99,100], and learn about physicians and hospitals to inform their choices [101]. Social media has become a primary information source for the public with varying information needs [1,102].

However, many studies raised concerns of social media information quality and its impact on individual health outcomes [103]. Indeed, empirical studies provided grounds for the concern. Sallam et al [104] found that people who relied on social media for health information had a lower COVID-19 knowledge score compared with those who relied on doctors and scientific journals for information. In addition, people who sought information on social media did not confirm its accuracy with health professionals, and health professionals disagreed with the information found on social media on 36.7% of occasions when people checked the information accuracy with health professionals [105]. This evidence calls for efforts to reduce misinformation on social media and improve the ability of the public to find reliable information sources on social media.

Exchange Social Support in Online Communities

Social support includes informational support, emotional support, esteem support, network support, and tangible support [106]. Except for tangible support, other types of social support were widely found in interactions in the online health community.

People with certain health concerns can receive and provide informational support content. They may read the experience of others diagnosed with the same illness, share details of their own medication and diagnosis [107,108], share and access medical knowledge such as treatment options and self-care activities [109], learn about health laws [110], and solicit recommendations of doctors and hospitals [99]. They may also proactively request health information and advice [111] and validate health information with others [112].

Emotional and esteem support, expressions that convey love and caring (emotional support) and respect and confidence in ability (esteem support), were considered as one category [106]. People with health concerns may express their disease feelings to mitigate their stress and appeal for sympathy [113], offer

XSL•FO
RenderX

hope and a sense of empowerment to each other [114], and encourage peers' engagement in medical care [115].

Network support content refers to expressions that communicate belonging to a group of people who have the same concerns or interest [106]: individual users use social media to build community with people who have similar health concerns or health goals [116], foster a group identity [116], enhance a sense of belonging [117], achieve shared health goals together [118], get connected with peer patients [119], and get involved in social events such as swimming lessons with peers [120]. Network support also empowered people with health concerns to influence health awareness at the societal level such as raising awareness of queer and trans issues [121].

People with health concerns may benefit from participating in online communities in several ways. First, they do not have to reveal personal identity and thus are free from the risk of being stigmatized when participating in online communities [107]. Second, online communities as an information channel supplement offline communication channels in terms of content and time. Peers may provide more information on treatment options not raised by health professionals because of time constraints [109]. Also, people can get answers and emotional support from online communities 24/7, while other social networks may not be able to provide such timely support [122]. Third, people have a sense of control over the help-seeking process when participating in online communities, which may fulfill patients' need of autonomy and reduce treatment avoidance [123]. Last, engaging in online communities may benefit health outcomes; studies found that perceived social support predicted subjective well-being [124].

Track and Share Health Statuses or Activities

Social media enables users to track and share their health statuses or activities and view those of others in the community. The public can document and share their illness experiences such as a breast cancer journey [125] or experience with chronic disease [126], achievements in health behavior change [127], and fitness activities [128] such as cycling and walking [129]. By self-tracking and sharing their physical activities, users can build networks with other users and discuss their activity performances [129]. Writing about and sharing their illness experiences may help patients cope with illness [125] and achieve health goals [126]. Other users may also benefit from such sharing as they can easily find peers who have similar experience to get advice on lifestyle changes and discuss treatment options [130].

Discussion

We reviewed articles published between 2006 and 2020 and identified 10 uses of social media for health purposes. The review found new social media uses for health purposes that emerged after 2013 as well as research gaps that need future research. The following sections will discuss the new uses and gaps.

New Social Media Uses for Health Purposes

Several new usages have emerged since 2013 compared with the usages summarized by Moorhead et al [12]. First, recent studies have begun to explore social media use for health purposes beyond the scope of patients, the general public, and health professionals. Studies have paid attention to social media use by health institutions including government agencies such as US federal, state, and local health departments [3,28,131] and nongovernmental health organizations [62]. Social media use identified in Moorhead et al [12] mostly benefits individuals who are experiencing illness or with the need for health information, while the use of social media by government agencies may benefit the whole society, especially during disease outbreaks. These articles also mentioned the limitations of government agencies using social media for health purposes such as lack of conversation with the public [131] and a disconnect between the health-related content generated by government agencies on social media and the actual health conditions that affect the population [2].

Second, studies in recent years have started to pay attention to social media use by health professionals and researchers to advance their practices and research. Social media is not only a communication channel that facilitates their conversation with patients [85] but also a tool for professional development [132] and health-related research [75]. Social media can provide opportunities for professional development from job searches to learning about trending research findings [77,80]. In addition, although studies have long been using social media for recruitment as a research tool, research efforts have been made to investigate the effectiveness of using social media for recruitment in recent years (ie, 2019-2020) and found social media promising as a recruitment tool [75].

Third, studies in recent years have found that social media can be used for social mobilization. Social media provides new opportunities for nongovernmental organizations to build community, mobilize resources, and extend discourse about sensitive issues beyond mainstream media coverage [47]. Although social mobilization was often initiated by nonprofit health organizations, people with heath concerns can form a community and collectively use social media to raise awareness of a disease or minority population. For example, patients with brain injuries can use social media to increase awareness of brain injury in society [133]. Social media was also used by LGBTQIA+ youth in the United States to increase the visibility of that population's voice and promote civic engagement on public issues related to the population [121].

Last, as social media incorporates more functions such as sending reminders, registering for events, and linking payment methods, social media becomes useful in facilitating offline health-related services and events such as making appointments and providing visiting guides [90]. In this sense, social media not only serves as a communication tool but also a tool for medical service and administration. However, this emerging usage may not be applicable to all social media platforms, as different platforms may incorporate functions not suitable for administrative uses. Moreover, this usage poses challenges to health organizations as they need to concern about the privacy issue and the need of additional workforces to manage their social media accounts [90].

XSL•FO
RenderX

Research Gaps in Social Media Uses for Health Purposes

There are several gaps existing in this research on using social media for health purposes. First, future studies should explore and test strategies to motivate engagement based on evidence of audience segmentation. There have been research efforts devoted to examining the audience segments on social media based on network characteristics in a health campaign [58], the extent to which users produce and consume health-related user-generated content [134], and users' motivation for participating in an online community [135]. Although these studies on audience segmentation provide implications on strategies to engage different types of audiences, we need more studies to formally test the effectiveness of strategies derived from audience segmentation research.

In addition, future research should develop a dynamic audience segment–detecting tool that can monitor audience characteristics and integrate these characteristics to predict audience segments on social media. Multiple factors may collaboratively determine audience types such as offline support [134], sentiment about the health issue [58], illness stage, need for support [135], and availability of spare time [136]. Some of these factors may change over time resulting in a transitioning of audience types [135]. As a result, changes in social media users' health status and social environments may lead to changes in their engagement pattern on social media. By predicting the change in user engagement patterns, practitioners can develop tailored engagement strategies for different types of social media users and adapt the strategies according to individual changes as an intervention goes on.

Second, more studies are needed exploring the relative impact of different applications of social media in health interventions on health outcomes, a research gap identified by Moorhead et al [12] that remains unresolved. Although the general impact of integrating social media in health interventions on promoting health knowledge and behavior is promising, it is unclear if different uses of social media in health interventions have different impacts. Social media can be integrated into an intervention in multiple ways and used for peer discussion with or without discussion leaders [137], conversations between participants and health professionals [138], delivering real-time and adaptive intervention messages [139] or regular intervention messages [45], delivering information in texts or more attractive formats such as video and infographics [138], intervention monitors to acknowledge achievements in participant health behavior changes [51], or participants to share their achievements with each other [127], etc. Using social media in these different manners may result in different effects on participant health behaviors and outcomes. For example, Garett et al [140] found that participants in the intervention group (ie, a Facebook group with peer leaders) acquired more knowledge about sexual health services and felt more comfortable with connecting with peers than those in the control group (ie, a Facebook group without peer leaders). Experimental studies and meta-analyses may be helpful to examine if and how using social media in an intervention in different ways affect health behavior change differently.

Third, social media enables users with similar health issues to develop health-related identities in a community. However, under which conditions identity building leads to positive or negative health behaviors and health outcomes is unclear [141]. Although developing identities may foster mutual support and decrease a sense of isolation [116], building an identity with a certain patient group can reinforce negative health behaviors such as disordered eating behaviors [142], and being part of a healthy community may pose pressures for users to consistently present an optimal identity as a healthy role model under persistent self- and community surveillance [128]. Researchers should explore approaches to using the identity-building function of social media to improve health outcomes and avoid negative impacts.

Last, privacy concerns were raised in articles on using social media for various health purposes including disease surveillance [21], health interventions [60], participation in online communities [143], and professionals communicating with colleagues and patients [76]. Researchers have consistently called for efforts to address privacy concerns related to using social media for health purposes [12,61,144]. Indeed, research effort has been made such as developing models to filter patients and caregivers and match them with trusted peer patients and caregivers in a privacy-preserving way [85] and developing social media privacy guide for health professionals [145]. However, gaps still exist in research and practice such as lacking an official guideline about privacy issues related to using social media in health research recruitment [73,146], lacking an approach to guarantee online informed consent [140], and researchers and potential participants lacking the awareness of the privacy risks of social media research recruitment [146]. Future studies should explore research methods and form practice guidelines that can address privacy issues associated with using social media for health research, intervention, and patient-doctor interaction.

Limitations

The review has several limitations. First, the review does not include conference papers, dissertations, and grey literature. However, given the large number of peer-reviewed journal articles included in the review and overlaps on social media uses identified in the articles, we expect that there will not be many other social media uses left unidentified by the review.

Second, we only searched keywords in abstracts instead of full texts. This may render the review missing articles that mentioned social media uses for health purposes only in the main text and not in the abstract. Indeed, some articles were not included in this review but were included in previous reviews; all articles included in this study were published between 2016 and 2020, while the previous review conducted by Moorhead and colleagues [12] included articles published before 2013. However, although the scope of this review may be limited due to the restrictive search strategy, the uses identified in the review cover most social media uses found by Moorhead and colleagues [12], meaning this study provides a comprehensive overview of the social media uses identified in existing literature. In addition, this review contributes to the literature by identifying social media uses for health purposes emerging after 2013.

XSL•FO
RenderX

Conclusions

This review summarized 10 social media uses for health purposes identified in previous literature and categorized the social media uses in terms of user types. Public health practitioners and organizations may use the summary as a starting point to explore applying social media in their daily practice. In addition, the review contributes to the literature by identifying research gaps in social media use for health purposes, providing guidance for future research in this area.

Conflicts of Interest

None declared.

References

1. Zhong Y, Liu W, Lee T, Zhao H, Ji J. Risk perception, knowledge, information sources and emotional states among COVID-19 patients in Wuhan, China. Nurs Outlook 2020 Aug 22;69(1):13-21 [FREE Full text] [doi: 10.1016/j.outlook.2020.08.005] [Medline: 32980153]
2. Braun LA, Zomorodbakhsch B, Keinki C, Huebner J. Information needs, communication and usage of social media by cancer patients and their relatives. J Cancer Res Clin Oncol 2019 Jul;145(7):1865-1875 [FREE Full text] [doi: 10.1007/s00432-019-02929-9] [Medline: 31123824]
3. Jha A, Lin L, Savoia E. The use of social media by state health departments in the US: analyzing health communication through Facebook. J Community Health 2016 Feb;41(1):174-179. [doi: 10.1007/s10900-015-0083-4] [Medline: 26318742]
4. Ranschaert ER, Van Ooijen PMA, McGinty GB, Parizel PM. Radiologists' usage of social media: results of the RANSOM survey. J Digit Imaging 2016 Aug;29(4):443-449 [FREE Full text] [doi: 10.1007/s10278-016-9865-1] [Medline: 26847202]
5. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horizons 2010 Jan;53(1):59-68 [FREE Full text] [doi: 10.1016/j.bushor.2009.09.003]
6. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
7. Yang Q. Are social networking sites making health behavior change interventions more effective? A meta-analytic review. J Health Commun 2017 Mar;22(3):223-233. [doi: 10.1080/10810730.2016.1271065] [Medline: 28248623]
8. Shi J, Poorisat T, Salmon CT. The use of social networking sites (SNSs) in health communication campaigns: review and recommendations. Health Commun 2018 Jan;33(1):49-56 [FREE Full text] [doi: 10.1080/10410236.2016.1242035] [Medline: 27858464]
9. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. Acad Med 2013 Jun;88(6):893-901. [doi: 10.1097/ACM.0b013e31828ffc23] [Medline: 23619071]
10. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
11. 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]
12. 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]
13. Moher D, Liberati A, Tetzlaff J, Altman DG. 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]
14. Krippendorff K. Reliability in content analysis: some common misconceptions and recommendations. Human Comm Res 2004 Jul 01;30(3):411-433 [FREE Full text] [doi: 10.1093/hcr/30.3.411]
15. 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]
16. Thorstad R, Wolff P. Predicting future mental illness from social media: a big-data approach. Behav Res Methods 2019 Aug;51(4):1586-1600 [FREE Full text] [doi: 10.3758/s13428-019-01235-z] [Medline: 31037606]
17. Edo-Osagie O, Smith G, Lake I, Edeghere O, De La Iglesia B. Twitter mining using semi-supervised classification for relevance filtering in syndromic surveillance. PLoS One 2019;14(7):e0210689 [FREE Full text] [doi: 10.1371/journal.pone.0210689] [Medline: 31318885]
18. Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, et al. Proactive Suicide Prevention Online (PSPO): machine identification and crisis management for chinese social media users with suicidal thoughts and behaviors. J Med Internet Res 2019 May 08;21(5):e11705 [FREE Full text] [doi: 10.2196/11705] [Medline: 31344675]

XSL•FO
RenderX

19. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
20. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. PLoS One 2020;15(11):e0241957 [FREE Full text] [doi: 10.1371/journal.pone.0241957] [Medline: 33170889]
21. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. J Biomed Inform 2020 Aug;108:103500 [FREE Full text] [doi: 10.1016/j.jbi.2020.103500] [Medline: 32622833]
22. Zhang J, Chen Y, Zhao Y, Wolfram D, Ma F. Public health and social media: a study of Zika virus‐related posts on Yahoo! Answers. J Assoc Inf Sci Technol 2019 May 17;71(3):282-299 [FREE Full text] [doi: 10.1002/asi.24245]
23. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social network analysis of COVID-19 sentiments: application of artificial intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
24. Doogan C, Buntine W, Linger H, Brunt S. Public perceptions and attitudes toward COVID-19 nonpharmaceutical interventions across six countries: a topic modeling analysis of twitter data. J Med Internet Res 2020 Sep 03;22(9):e21419 [FREE Full text] [doi: 10.2196/21419] [Medline: 32784190]
25. Zhang H, Wheldon C, Dunn AG, Tao C, Huo J, Zhang R, et al. Mining Twitter to assess the determinants of health behavior toward human papillomavirus vaccination in the United States. J Am Med Inform Assoc 2020 Feb 01;27(2):225-235 [FREE Full text] [doi: 10.1093/jamia/ocz191] [Medline: 31711186]
26. Eden OB, Lilleyman J, Shaw MP, Richards S, Peto J. Medical Research Council Childhood Leukaemia Trial VIII compared with trials II-VII: lessons for future management. Haematol Blood Transfus 1987;30:448-455 [FREE Full text] [doi: 10.1007/978-3-642-71213-5_79] [Medline: 3305212]
27. Hatchard JL, Quariguasi Frota Neto J, Vasilakis C, Evans-Reeves KA. Tweeting about public health policy: social media response to the UK Government's announcement of a Parliamentary vote on draft standardised packaging regulations. PLoS One 2019;14(2):e0211758 [FREE Full text] [doi: 10.1371/journal.pone.0211758] [Medline: 30807582]
28. Jin Y, Austin L, Vijaykumar S, Jun H, Nowak G. Communicating about infectious disease threats: insights from public health information officers. Public Relations Rev 2019 Mar;45(1):167-177 [FREE Full text] [doi: 10.1016/j.pubrev.2018.12.003]
29. Vandormael A, Adam M, Greuel M, Bärnighausen T. An entertainment-education approach to prevent COVID-19 spread: study protocol for a multi-site randomized controlled trial. Trials 2020 Dec 15;21(1):1025 [FREE Full text] [doi: 10.1186/s13063-020-04942-7] [Medline: 33323130]
30. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: how to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
31. Sutton J, Renshaw SL, Butts CT. COVID-19: retransmission of official communications in an emerging pandemic. PLoS ONE 2020 Sep 16;15(9):e0238491 [FREE Full text] [doi: 10.1371/journal.pone.0238491]
32. Gatewood J, Monks SL, Singletary CR, Vidrascu E, Moore JB. Social media in public health: strategies to distill, package, and disseminate public health research. J Public Health Manag Pract 2020;26(5):489-492 [FREE Full text] [doi: 10.1097/PHH.0000000000001096] [Medline: 32732723]
33. McInnes C, Hornmoen H. Add Twitter and Stir: the use of Twitter by public authorities in Norway and UK during the 2014-15 Ebola outbreak. Observatorio 2018;12(2):23-46 [FREE Full text] [doi: 10.15847/obsobs12220181173]
34. Harrison D, Wilding J, Bowman A, Fuller A, Nicholls SG, Pound CM, et al. Using YouTube to disseminate effective vaccination pain treatment for babies. PLoS One 2016;11(10):e0164123 [FREE Full text] [doi: 10.1371/journal.pone.0164123] [Medline: 27695054]
35. Patrick MD, Stukus DR, Nuss KE. Using podcasts to deliver pediatric educational content: development and reach of PediaCast CME. Digit Health 2019;5:2055207619834842 [FREE Full text] [doi: 10.1177/2055207619834842] [Medline: 30834137]
36. Grimes DR. EMBO Rep 2020 Nov 05;21(11):e51819 [FREE Full text] [doi: 10.15252/embr.202051819] [Medline: 33155436]
37. Xie T, Tan T, Li J. An extensive search trends-based analysis of public attention on social media in the early outbreak of COVID-19 in China. Risk Manag Healthc Policy 2020;13:1353-1364 [FREE Full text] [doi: 10.2147/RMHP.S257473] [Medline: 32943953]
38. Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC Public Health 2019 Oct 23;19(1):1348 [FREE Full text] [doi: 10.1186/s12889-019-7659-3] [Medline: 31640660]
39. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
40. Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. Social Media can have an impact on how we manage and investigate the COVID-19 pandemic. J Clin Epidemiol 2020 Jun 27:127-201 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.06.028] [Medline: 32603686]

XSL•FO
RenderX

41. Southwell BG, Wood JL, Navar AM. Roles for health care professionals in addressing patient-held misinformation beyond fact correction. Am J Public Health 2020 Oct;110(S3):S288-S289 [FREE Full text] [doi: 10.2105/AJPH.2020.305729] [Medline: 33001723]
42. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch GS. Correcting misinformation by health organizations during measles outbreaks: a controlled experiment. PLoS One 2018;13(12):e0209505 [FREE Full text] [doi: 10.1371/journal.pone.0209505] [Medline: 30566485]
43. Diddi P, Lundy LK. Organizational Twitter use: content analysis of tweets during breast cancer awareness month. J Health Commun 2017 Dec;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
44. Li H, Xue L, Tucker JD, Wei C, Durvasula M, Hu W, et al. Condom use peer norms and self-efficacy as mediators between community engagement and condom use among Chinese men who have sex with men. BMC Public Health 2017 Aug 07;17(1):641 [FREE Full text] [doi: 10.1186/s12889-017-4662-4] [Medline: 28784172]
45. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
46. Payton F. Cultures of participation—for students, by students. Info Systems J 2015 Aug 14;26(4):319-338 [FREE Full text] [doi: 10.1111/isj.12086]
47. Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. J Health Commun 2016 May;21(5):517-526. [doi: 10.1080/10810730.2015.1103332] [Medline: 27115179]
48. Park H, Reber BH, Chon M. Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement. J Health Commun 2016;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
49. Yu Y, Li Y, Li T, Xi S, Xiao X, Xiao S, et al. New path to recovery and well-being: cross-sectional study on Wechat use and endorsement of Wechat-based mHealth among people living with schizophrenia in China. J Med Internet Res 2020 Sep 18;22(9):e18663 [FREE Full text] [doi: 10.2196/18663] [Medline: 32945774]
50. Fernandez-Luque L, Singh M, Ofli F, Mejova YA, Weber I, Aupetit M, et al. Implementing 360° Quantified Self for childhood obesity: feasibility study and experiences from a weight loss camp in Qatar. BMC Med Inform Decis Mak 2017 Apr 13;17(1):37 [FREE Full text] [doi: 10.1186/s12911-017-0432-6] [Medline: 28403865]
51. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12):1. [doi: 10.1002/pbc.26660] [Medline: 28618158]
52. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to health: a novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2016 Jun 16;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
53. Fang Y, Ma Y, Mo D, Zhang S, Xiang M, Zhang Z. Methodology of an exercise intervention program using social incentives and gamification for obese children. BMC Public Health 2019 Jun 03;19(1):686 [FREE Full text] [doi: 10.1186/s12889-019-6992-x] [Medline: 31159776]
54. Pagkas-Bather J, Young LE, Chen Y, Schneider JA. Social network interventions for HIV transmission elimination. Curr HIV/AIDS Rep 2020 Oct;17(5):450-457 [FREE Full text] [doi: 10.1007/s11904-020-00524-z] [Medline: 32720253]
55. Thornber K, Huso D, Rahman MM, Biswas H, Rahman MH, Brum E, et al. Raising awareness of antimicrobial resistance in rural aquaculture practice in Bangladesh through digital communications: a pilot study. Glob Health Action 2019;12(sup1):1734735 [FREE Full text] [doi: 10.1080/16549716.2020.1734735] [Medline: 32153258]
56. Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B. Review of evaluation metrics used in digital and traditional tobacco control campaigns. J Med Internet Res 2020 Aug 11;22(8):e17432 [FREE Full text] [doi: 10.2196/17432] [Medline: 32348272]
57. Moreno MA, D'Angelo J. Social media intervention design: applying an affordances framework. J Med Internet Res 2019 Mar 26;21(3):e11014 [FREE Full text] [doi: 10.2196/11014] [Medline: 30912754]
58. Chu K, Allem J, Unger JB, Cruz TB, Akbarpour M, Kirkpatrick MG. Strategies to find audience segments on Twitter for e-cigarette education campaigns. Addict Behav 2019 Apr;91:222-226 [FREE Full text] [doi: 10.1016/j.addbeh.2018.11.015] [Medline: 30497815]
59. Xiao S, Li T, Zhou W, Shen M, Yu Y. WeChat-based mHealth intention and preferences among people living with schizophrenia. PeerJ 2020;8:e10550 [FREE Full text] [doi: 10.7717/peerj.10550] [Medline: 33362979]
60. Ronen K, Grant E, Copley C, Batista T, Guthrie BL. Peer group focused ehealth strategies to promote HIV prevention, testing, and care engagement. Curr HIV/AIDS Rep 2020 Oct;17(5):557-576 [FREE Full text] [doi: 10.1007/s11904-020-00527-w] [Medline: 32794071]
61. Bakken S, Marden S, Arteaga SS, Grossman L, Keselman A, Creber RM, et al. Behavioral interventions using consumer information technology as tools to advance health equity. Am J Public Health 2019 Jan;109(S1):S79-S85. [doi: 10.2105/AJPH.2018.304646] [Medline: 30699018]
62. Fu J, Zhang R. NGOs' HIV/AIDS discourse on social media and websites: technology affordances and strategic communication across media platforms. Int J Comm 2019;13:25.

XSL•FO
RenderX

63. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Dec;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
64. Ren J, Raghupathi V, Raghupathi W. Understanding the dimensions of medical crowdfunding: a visual analytics approach. J Med Internet Res 2020 Jul 03;22(7):e18813 [FREE Full text] [doi: 10.2196/18813] [Medline: 32618573]
65. Young MJ, Scheinberg E. The rise of crowdfunding for medical care: promises and perils. JAMA 2017 Apr 25;317(16):1623-1624 [FREE Full text] [doi: 10.1001/jama.2017.3078] [Medline: 28334321]
66. He S, Ojo A, Beckman AL, 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]
67. Guidry J, Zhang Y, Jin Y, Parrish C. Portrayals of depression on Pinterest and why public relations practitioners should care. Public Relations Rev 2016 Mar;42(1):232-236 [FREE Full text] [doi: 10.1016/j.pubrev.2015.09.002]
68. Jiang T, Osadchiy V, Mills JN, Eleswarapu SV. Is it all in my head? Self-reported psychogenic erectile dysfunction and depression are common among young men seeking advice on social media. Urology 2020 May 11;142:133-140. [doi: 10.1016/j.urology.2020.04.100] [Medline: 32437776]
69. Oser TK, Minnehan KA, Wong G, Parascando J, McGinley E, Radico J, et al. Using social media to broaden understanding of the barriers and facilitators to exercise in adults with type 1 diabetes. J Diabetes Sci Technol 2019 May;13(3):457-465. [doi: 10.1177/1932296819835787] [Medline: 30862185]
70. Erikainen S, Pickersgill M, Cunningham-Burley S, Chan S. Patienthood and participation in the digital era. Digit Health 2019;5:2055207619845546 [FREE Full text] [doi: 10.1177/2055207619845546] [Medline: 31041112]
71. Koutrolou-Sotiropoulou P, Lima FV, Stergiopoulos K. Quality of life in survivors of peripartum cardiomyopathy. Am J Cardiol 2016 Jul 15;118(2):258-263 [FREE Full text] [doi: 10.1016/j.amjcard.2016.04.040] [Medline: 27239023]
72. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med 2019 Jan 16;17(1):11 [FREE Full text] [doi: 10.1186/s12916-018-1247-8] [Medline: 30646913]
73. Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, et al. Social media recruitment for mental health research: a systematic review. Compr Psychiatry 2020 Nov;103:152197 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152197] [Medline: 32992073]
74. 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. [doi: 10.1007/s10900-019-00677-y] [Medline: 31102117]
75. Shaver LG, Khawer A, Yi Y, Aubrey-Bassler K, Etchegary H, Roebothan B, et al. Using Facebook advertising to recruit representative samples: feasibility assessment of a cross-sectional survey. J Med Internet Res 2019 Aug 19;21(8):e14021 [FREE Full text] [doi: 10.2196/14021] [Medline: 31429409]
76. Cherrez-Ojeda I, Vanegas E, Felix M, Mata VL, Jiménez FM, Sanchez M, et al. Frequency of use, perceptions and barriers of information and communication technologies among Latin American physicians: an Ecuadorian cross-sectional study. J Multidiscip Healthc 2020;13:259-269 [FREE Full text] [doi: 10.2147/JMDH.S246253] [Medline: 32214821]
77. Mascia D, Rinninella E, Pennacchio NW, Cerrito L, Gasbarrini A. It's how we communicate! Exploring face-to-face versus electronic communication networks in multidisciplinary teams. Health Care Manage Rev 2019 Apr 17:1 [FREE Full text] [doi: 10.1097/HMR.0000000000000246] [Medline: 31008805]
78. Yüce, Adalı E, Kanmaz B. An analysis of YouTube videos as educational resources for dental practitioners to prevent the spread of COVID-19. Ir J Med Sci 2021 Mar;190(1):19-26 [FREE Full text] [doi: 10.1007/s11845-020-02312-5] [Medline: 32700083]
79. Bamat NA, Manley BJ, Harer MW, Roland D. Social media for pediatric research: what, who, why, and #? Pediatr Res 2018 Nov;84(5):597-599 [FREE Full text] [doi: 10.1038/s41390-018-0140-7] [Medline: 30140067]
80. Helm J, Jones RM. Practice paper of the academy of nutrition and dietetics: social media and the dietetics practitioner: opportunities, challenges, and best practices. J Acad Nutr Diet 2016 Nov;116(11):1825-1835. [doi: 10.1016/j.jand.2016.09.003] [Medline: 27788767]
81. Loeb S, Carrick T, Frey C, Titus T. Increasing social media use in urology: 2017 American Urological Association Survey. Eur Urol Focus 2020 May 15;6(3):605-608 [FREE Full text] [doi: 10.1016/j.euf.2019.07.004] [Medline: 31351900]
82. Pizzuti AG, Patel KH, McCreary EK, Heil E, Bland CM, Chinaeke E, et al. Healthcare practitioners' views of social media as an educational resource. PLoS One 2020;15(2):e0228372 [FREE Full text] [doi: 10.1371/journal.pone.0228372] [Medline: 32027686]
83. Borgmann H, DeWitt S, Tsaur I, Haferkamp A, Loeb S. Novel survey disseminated through Twitter supports its utility for networking, disseminating research, advocacy, clinical practice and other professional goals. Can Urol Assoc J 2015;9(9-10):E713-E717 [FREE Full text] [doi: 10.5489/cuaj.3014] [Medline: 26664662]
84. Birnbaum ML, Rizvi AF, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2015 Mar 23;11(4):290-295. [doi: 10.1111/eip.12237] [Medline: 25808317]
85. Tang W, Ren J, Zhang Y. Enabling trusted and privacy-preserving healthcare services in social media health networks. IEEE Trans Multimedia 2019 Mar;21(3):579-590 [FREE Full text] [doi: 10.1109/tmm.2018.2889934]

XSL•FO
RenderX

86. Benetoli A, Chen TF, Schaefer M, Chaar B, Aslani P. Do pharmacists use social media for patient care? Int J Clin Pharm 2017 Apr;39(2):364-372 [FREE Full text] [doi: 10.1007/s11096-017-0444-4] [Medline: 28299527]
87. Hermansyah A, Sukorini AI, Asmani F, Suwito KA, Rahayu TP. The contemporary role and potential of pharmacist contribution for community health using social media. J Basic Clin Physiol Pharmacol 2019 Dec 04;30(6):1 [FREE Full text] [doi: 10.1515/jbcpp-2019-0329] [Medline: 31800395]
88. Senthilkumaran S, Benita F, Balamurugan N, Thirumalaikolundusubramanian P. Social media for health care education—can it make a difference? Am J Emerg Med 2018 Jan;36(1):150-151 [FREE Full text] [doi: 10.1016/j.ajem.2017.07.027] [Medline: 28709715]
89. Liu Q, Liu X, Guo X. The effects of participating in a physician-driven online health community in managing chronic disease: evidence from two natural experiments. MISQ 2020 Jan 01;44(1):391-419 [FREE Full text] [doi: 10.25300/misq/2020/15102]
90. Shen L, Wang S, Chen W, Fu Q, Evans R, Lan F, et al. Understanding the function constitution and influence factors on communication for the wechat official account of top tertiary hospitals in china: cross-sectional study. J Med Internet Res 2019 Dec 09;21(12):e13025 [FREE Full text] [doi: 10.2196/13025] [Medline: 31815674]
91. Peluchette JV, Karl KA, Coustasse A. Physicians, patients, and Facebook: Could you? Would you? Should you? Health Mark Q 2016;33(2):112-126 [FREE Full text] [doi: 10.1080/07359683.2016.1166811] [Medline: 27295007]
92. Kenny P, Johnson IG. Social media use, attitudes, behaviours and perceptions of online professionalism amongst dental students. Br Dent J 2016 Nov 18;221(10):651-655. [doi: 10.1038/sj.bdj.2016.864] [Medline: 27857111]
93. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. Health Info Libr J 2019 Mar;36(1):60-72 [FREE Full text] [doi: 10.1111/hir.12247] [Medline: 30663232]
94. Raggatt M, Wright CJC, Carrotte E, Jenkinson R, Mulgrew K, Prichard I, et al. "I aspire to look and feel healthy like the posts convey": engagement with fitness inspiration on social media and perceptions of its influence on health and wellbeing. BMC Public Health 2018 Aug 10;18(1):1002 [FREE Full text] [doi: 10.1186/s12889-018-5930-7] [Medline: 30097034]
95. Gomez-Lopez IN, Clarke P, Hill AB, Romero DM, Goodspeed R, Berrocal VJ, et al. Using social media to identify sources of healthy food in urban neighborhoods. J Urban Health 2017 Jun;94(3):429-436 [FREE Full text] [doi: 10.1007/s11524-017-0154-1] [Medline: 28455606]
96. Lambert M, Chivers P, Farringdon F. In their own words: a qualitative study exploring influences on the food choices of university students. Health Promot J Austr 2019 Jan;30(1):66-75 [FREE Full text] [doi: 10.1002/hpja.180] [Medline: 29892990]
97. Zhu C, Zeng R, Zhang W, Evans R, He R. Pregnancy-related information seeking and sharing in the social media era among expectant mothers: qualitative study. J Med Internet Res 2019 Dec 04;21(12):e13694 [FREE Full text] [doi: 10.2196/13694] [Medline: 31799939]
98. Schwartz B. How often do patients turn to social media for STD second opinions? Contemporary OB/GYN 2019;64(12):18.
99. Wu T, He Z, Zhang D. Impact of communicating with doctors via social media on consumers' e-health literacy and healthy behaviors in China. Inquiry 2020;57:46958020971188 [FREE Full text] [doi: 10.1177/0046958020971188] [Medline: 33238788]
100. Ahmed N, Quinn SC, Hancock GR, Freimuth VS, Jamison A. Social media use and influenza vaccine uptake among White and African American adults. Vaccine 2018 Nov 26;36(49):7556-7561 [FREE Full text] [doi: 10.1016/j.vaccine.2018.10.049] [Medline: 30389192]
101. Tengilimoglu D, Sarp N, Yar CE, Bektaş M, Hidir MN, Korkmaz E. The consumers' social media use in choosing physicians and hospitals: the case study of the province of Izmir. Int J Health Plann Manage 2017 Jan;32(1):19-35. [doi: 10.1002/hpm.2296] [Medline: 26198772]
102. Sagnia PIG, Gharoro EP, Isara AR. Adolescent-parent communication on sexual and reproductive health issues amongst secondary school students in Western Region 1 of The Gambia. Afr J Prim Health Care Fam Med 2020 Nov 04;12(1):e1-e7 [FREE Full text] [doi: 10.4102/phcfm.v12i1.2437] [Medline: 33181882]
103. Reidy C, Klonoff DC, Barnard-Kelly KD. Supporting good intentions with good evidence: how to increase the benefits of diabetes social media. J Diabetes Sci Technol 2019 Sep;13(5):974-978 [FREE Full text] [doi: 10.1177/1932296819850187] [Medline: 31096768]
104. Sallam M, Dababseh D, Yaseen A, Al-Haidar A, Taim D, Eid H, et al. COVID-19 misinformation: mere harmless delusions or much more? A knowledge and attitude cross-sectional study among the general public residing in Jordan. PLoS One 2020;15(12):e0243264 [FREE Full text] [doi: 10.1371/journal.pone.0243264] [Medline: 33270783]
105. Crilly P, Jair S, Mahmood Z, Moin Khan A, Munir A, Osei-Bediako I, et al. Public views of different sources of health advice: pharmacists, social media and mobile health applications. Int J Pharm Pract 2019 Mar;27(1):88-95 [FREE Full text] [doi: 10.1111/ijpp.12448] [Medline: 29732649]
106. Cutrona C, Suhr J. Controllability of stressful events and satisfaction with spouse support behaviors. Comm Res 2016 Jun 30;19(2):154-174 [FREE Full text] [doi: 10.1177/009365092019002002]

XSL•FO
RenderX

107. 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]
108. Årsand E, Bradway M, Gabarron E. What are diabetes patients versus health care personnel discussing on social media? J Diabetes Sci Technol 2019 Mar;13(2):198-205 [FREE Full text] [doi: 10.1177/1932296818821649] [Medline: 30636445]
109. Rupert DJ, Gard Read J, Amoozegar JB, Moultrie RR, Taylor OM, O'Donoghue AC, et al. Peer-generated health information: the role of online communities in patient and caregiver health decisions. J Health Commun 2016 Nov;21(11):1187-1197 [FREE Full text] [doi: 10.1080/10810730.2016.1237592] [Medline: 27805496]
110. Xu Y, Burleson B. Effects of sex, culture, and support type on perceptions of spousal social support: an assessment of the "support gap" hypothesis in early marriage. Human Comm Res 2001 Oct;27(4):535-566 [FREE Full text] [doi: 10.1111/j.1468-2958.2001.tb00792.x]
111. Yeo TED, Chu TH. Sharing "sex secrets" on Facebook: a content analysis of youth peer communication and advice exchange on social media about sexual health and intimate relations. J Health Commun 2017 Sep;22(9):753-762 [FREE Full text] [doi: 10.1080/10810730.2017.1347217] [Medline: 28796578]
112. McCarthy R, Byrne G, Brettle A, Choucri L, Ormandy P, Chatwin J. Midwife-moderated social media groups as a validated information source for women during pregnancy. Midwifery 2020 Sep;88:102710. [doi: 10.1016/j.midw.2020.102710] [Medline: 32485501]
113. 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):3561 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
114. Prescott J, Rathbone AL, Brown G. Online peer to peer support: qualitative analysis of UK and US open mental health Facebook groups. Digit Health 2020;6:2055207620979209 [FREE Full text] [doi: 10.1177/2055207620979209] [Medline: 33354335]
115. Platt JR, Brady RR. #BCSM and #breastcancer: contemporary cancer-specific online social media communities. Breast J 2020 Apr;26(4):729-733. [doi: 10.1111/tbj.13576] [Medline: 31493301]
116. Byron P, Hunt J. 'That happened to me too': young people's informal knowledge of diverse genders and sexualities. Sex Education 2017 Feb 22;17(3):319-332 [FREE Full text] [doi: 10.1080/14681811.2017.1292899]
117. Zigron S, Bronstein J. "Help is where you find it": the role of weak ties networks as sources of information and support in virtual health communities. J Assoc Inf Sci Technol 2018 Nov 19;70(2):130-139 [FREE Full text] [doi: 10.1002/asi.24106]
118. Chaddha A, Jackson EA, Richardson CR, Franklin BA. Technology to help promote physical activity. Am J Cardiol 2017 Jan 01;119(1):149-152 [FREE Full text] [doi: 10.1016/j.amjcard.2016.09.025] [Medline: 27889045]
119. Lerman B, Lewis S, Lumley M, Grogan G, Hudson C, Johnson E. Teen depression groups on Facebook: a content analysis. J Adolesc Res 2016 Oct 22;32(6):719-741 [FREE Full text] [doi: 10.1177/0743558416673717]
120. Cole L, Kharwa Y, Khumalo N, Reinke J, Karrim S. Caregivers of school-aged children with autism: social media as a source of support. J Child Fam Stud 2017 Aug 31;26(12):3464-3475 [FREE Full text] [doi: 10.1007/s10826-017-0855-9]
121. Manduley A, Mertens A, Plante I, Sultana A. The role of social media in sex education: dispatches from queer, trans, and racialized communities. Feminism Psychol 2018 Feb 08;28(1):152-170 [FREE Full text] [doi: 10.1177/0959353517717751]
122. Clapton-Caputo E, Sweet L, Muller A. A qualitative study of expectations and experiences of women using a social media support group when exclusively expressing breastmilk to feed their infant. Women Birth 2020 Jul 13:1. [doi: 10.1016/j.wombi.2020.06.010] [Medline: 32674991]
123. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
124. Han X, Li B, Qu J, Zhu Q. Weibo friends with benefits for people live with HIV/AIDS? The implications of Weibo use for enacted social support, perceived social support and health outcomes. Soc Sci Med 2018 Aug;211:157-163 [FREE Full text] [doi: 10.1016/j.socscimed.2018.06.016] [Medline: 29940485]
125. Wan Adnan W, Ibrahim N, Haladin N, Wan Fakhruddin W. Illness experience sharing on personal blogs: a case study of malaysian breast cancer survivors. GEMA 2019 Nov 29;19(4):28-47 [FREE Full text] [doi: 10.17576/gema-2019-1904-02]
126. Song J, Xu P, Paradice DB. Health goal attainment of patients with chronic diseases in web-based patient communities: content and survival analysis. J Med Internet Res 2020 Sep 11;22(9):e19895 [FREE Full text] [doi: 10.2196/19895] [Medline: 32915152]
127. Armin J, Johnson T, Hingle M, Giacobbi P, Gordon JS. Development of a multi-behavioral mhealth app for women smokers. J Health Commun 2017 Feb;22(2):153-162. [doi: 10.1080/10810730.2016.1256454] [Medline: 28121240]
128. Kent R. Self-tracking health over time: from the use of instagram to perform optimal health to the protective shield of the digital detox. Soc Media Soc 2020 Sep 18;6(3):205630512094069 [FREE Full text] [doi: 10.1177/2056305120940694]
129. Carter S, Green J, Speed E. Digital technologies and the biomedicalisation of everyday activities: the case of walking and cycling. Sociol Compass 2018 Mar 15;12(4):e12572 [FREE Full text] [doi: 10.1111/soc4.12572]
130. Al Walid M. Data analysis and visualization of continental cancer situation by twitter scraping. IJMECS 2019 Jul 08;11(7):23-31 [FREE Full text] [doi: 10.5815/ijmecs.2019.07.03]

XSL•FO
RenderX

131. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49 [FREE Full text] [doi: 10.1186/s12911-017-0447-z] [Medline: 28431582]
132. Carlquist E, Lee NE, Shalin SC, Goodman M, Gardner JM. Dermatopathology and social media: improving participation by increasing clarity. Arch Pathol Lab Med 2018 Jan;142(1):11 [FREE Full text] [doi: 10.5858/arpa.2017-0331-LE] [Medline: 29257927]
133. Brandhorst J, Jennings F. Fighting for funding: values advocacy and Planned Parenthood's right-to-life. Public Relations Rev 2016 Nov;42(4):723-733 [FREE Full text] [doi: 10.1016/j.pubrev.2016.06.002]
134. Fergie G, Hunt K, Hilton S. Social media as a space for support: young adults' perspectives on producing and consuming user-generated content about diabetes and mental health. Soc Sci Med 2016 Dec;170:46-54 [FREE Full text] [doi: 10.1016/j.socscimed.2016.10.006] [Medline: 27750067]
135. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas in online health communities. J Biomed Inform 2016 Oct;63:212-225. [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]
136. De Simoni A, Shah AT, Fulton O, Parkinson J, Sheikh A, Panzarasa P, et al. Superusers' engagement in asthma online communities: asynchronous web-based interview study. J Med Internet Res 2020 Jun 23;22(6):e18185 [FREE Full text] [doi: 10.2196/18185] [Medline: 32573463]
137. Zhao Y, Zhu X, Pérez AE, Zhang W, Shi A, Zhang Z, et al. MHealth approach to promote Oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146 [FREE Full text] [doi: 10.1186/s12889-018-6046-9] [Medline: 30261856]
138. Condran B, Gahagan J, Isfeld-Kiely H. A scoping review of social media as a platform for multi-level sexual health promotion interventions. Can J Human Sexuality 2017 Apr;26(1):26-37 [FREE Full text] [doi: 10.3138/cjhs.261-a1]
139. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb;42(1):67-83. [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
140. Garett R, Menacho L, Young SD. Ethical issues in using social media to deliver an HIV prevention intervention: results from the HOPE Peru study. Prev Sci 2017 Feb;18(2):225-232 [FREE Full text] [doi: 10.1007/s11121-016-0739-z] [Medline: 27933425]
141. O'Neill B. Towards an improved understanding of modern health information ecology. Soc Sci Med 2017 Jan;173:108-109 [FREE Full text] [doi: 10.1016/j.socscimed.2016.12.003] [Medline: 27951460]
142. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
143. Gow J, Moffatt C, Blackport J. Participation in patient support forums may put rare disease patient data at risk of re-identification. Orphanet J Rare Dis 2020 Aug 31;15(1):226 [FREE Full text] [doi: 10.1186/s13023-020-01497-3] [Medline: 32867839]
144. Wang S, Ding S, Xiong L. A new system for surveillance and digital contact tracing for COVID-19: spatiotemporal reporting over network and GPS. JMIR Mhealth Uhealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]
145. Hennessy CM, Smith CF, Greener S, Ferns G. Social media guidelines: a review for health professionals and faculty members. Clin Teach 2019 Oct;16(5):442-447. [doi: 10.1111/tct.13033] [Medline: 31144449]
146. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: A privacy-by-design framework for online recruitment. J Med Internet Res 2017;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]

Abbreviations

LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, asexual
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by G Eysenbach; submitted 22.01.20; peer-reviewed by H Korda, JR Bautista; comments to author 23.09.20; revised version received 29.01.21; accepted 11.04.21; published 12.05.21.

Please cite as:
Chen J, Wang Y
Social Media Use for Health Purposes: Systematic Review
J Med Internet Res 2021;23(5):e17917
URL: https://www.jmir.org/2021/5/e17917
doi:10.2196/17917
PMID:33978589

XSL•FO
RenderX

©Junhan Chen, Yuan Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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 Digital Interventions for Reducing Behavioral Risks of Cardiovascular Disease in Nonclinical Adult Populations: Systematic Review of Reviews

Natalie Gold[1,2], DPhil; Amy Yau[1,3], PhD; Benjamin Rigby[1,4], MA; Chris Dyke[1,5], MSc; Elizabeth Alice Remfry[1,6], BSc; Tim Chadborn[1], PhD

[1]Public Health England, London, United Kingdom

[2]Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, United Kingdom

[3]Centre for Diet and Activity Research, MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

[4]Department of Sociology, University of Durham, Durham, United Kingdom

[5]Department of Social Science, Institute of Education, University College London, London, United Kingdom

[6]Global Health Section, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Natalie Gold, DPhil
Public Health England
Wellington House
133-155 Waterloo Road
London, SE1 8UG
United Kingdom
Phone: 44 9767839044
Email: n.gold@lse.ac.uk

Abstract

Background: Digital health interventions are increasingly being used as a supplement or replacement for face-to-face services as a part of predictive prevention. They may be offered to those who are at high risk of cardiovascular disease and need to improve their diet, increase physical activity, stop smoking, or reduce alcohol consumption. Despite the popularity of these interventions, there is no overall summary and comparison of the effectiveness of different modes of delivery of a digital intervention to inform policy.

Objective: This review aims to summarize the effectiveness of digital interventions in improving behavioral and health outcomes related to physical activity, smoking, alcohol consumption, or diet in nonclinical adult populations and to identify the effectiveness of different modes of delivery of digital interventions.

Methods: We reviewed articles published in the English language between January 1, 2009, and February 25, 2019, that presented a systematic review with a narrative synthesis or meta-analysis of any study design examining digital intervention effectiveness; data related to adults (≥18 years) in high-income countries; and data on behavioral or health outcomes related to diet, physical activity, smoking, or alcohol, alone or in any combination. Any time frame or comparator was considered eligible. We searched MEDLINE, Embase, PsycINFO, Cochrane Reviews, and gray literature. The AMSTAR-2 tool was used to assess review confidence ratings.

Results: We found 92 reviews from the academic literature (47 with meta-analyses) and 2 gray literature items (1 with a meta-analysis). Digital interventions were typically more effective than no intervention, but the effect sizes were small. Evidence on the effectiveness of digital interventions compared with face-to-face interventions was mixed. Most trials reported that intent-to-treat analysis and attrition rates were often high. Studies with long follow-up periods were scarce. However, we found that digital interventions may be effective for up to 6 months after the end of the intervention but that the effects dissipated by 12 months. There were small positive effects of digital interventions on smoking cessation and alcohol reduction; possible effectiveness in combined diet and physical activity interventions; no effectiveness for interventions targeting physical activity alone, except for when interventions were delivered by mobile phone, which had medium-sized effects; and no effectiveness observed for interventions targeting diet alone. Mobile interventions were particularly effective. Internet-based interventions were generally effective.

XSL•FO
RenderX

Conclusions: Digital interventions have small positive effects on smoking, alcohol consumption, and in interventions that target a combination of diet and physical activity. Small effects may have been due to the low efficacy of treatment or due to nonadherence. In addition, our ability to make inferences from the literature we reviewed was limited as those interventions were heterogeneous, many reviews had critically low AMSTAR-2 ratings, analysis was typically intent-to-treat, and follow-up times were relatively short.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019126074; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=126074.

(J Med Internet Res 2021;23(5):e19688) doi:10.2196/19688

KEYWORDS

alcohol; behavior change; cardiovascular disease; diet; digital interventions; digital medicine; internet interventions; mHealth; mobile interventions; physical activity; smoking; tobacco; mobile phone

Introduction

Background

The National Health Service (NHS) Long Term Plan sets out the UK government's vision for preventing health problems and supporting the self-management of conditions [1]. A major target is cardiovascular disease (CVD), which causes 28% of all deaths in the United Kingdom and is the largest cause of premature death in deprived areas [2,3]. England's primary large-scale intervention for CVD prevention is the NHS Health Check program [1,2], which was introduced in 2009 [4]. It is one of the largest public health prevention programs in the world, with over 6 million people in England having a check between 2013 and 2018 [5]. The NHS Health Check is a CVD risk assessment, which should be offered every 5 years to all adults aged between 40 and 74 years with no pre-existing vascular condition. As a result, people with previously undiagnosed conditions can be put on a clinical pathway and those who are at risk of developing a condition can be offered lifestyle support and advice to manage their risk. In particular, cardiovascular risk can be reduced by modifying 4 types of behavior: diet [6], physical activity [7], smoking [8], and alcohol consumption [9].

A key pillar of the Long Term Plan is *predictive prevention*—the use of technology and digital tools to identify health risks, make early diagnoses, and support positive health behaviors of those most at need through targeted treatments [3]. Predictive prevention, including the NHS Health Check, involves risk communication and behavior change. Evidence shows that risk communication alone does not lead to behavior change [10-13]. Therefore, we need to support behavior change. Digital tools are an increasingly important part of that landscape, and digital behavior change interventions may be offered to people after their NHS Health Check to manage their risk by helping them modify their diet, physical activity, smoking, or alcohol consumption.

Digital tools may be used either to supplement face-to-face services or to replace them. Replacement is particularly germane, since there is anecdotal evidence that face-to-face services are increasingly being defunded. In addition, services may need to shift from face-to-face to digital in response to the COVID-19 pandemic, which occurred after we had completed the review. Providers may hope that digital tools will offer a low-cost solution, with the potential to reach more people than traditional face-to-face services; however, the research base needs to be evaluated to see if there is sufficient evidence [14].

Aims and Objectives

The first step is to establish whether digital interventions are effective. We also need to know which modes of delivery are most effective in order to allocate resources to develop the most promising digital tools or to know where research is needed, if the evidence base is lacking. In this systematic review of reviews, we aim to summarize the evidence on the effectiveness of digital interventions in improving dietary, physical activity, smoking, and alcohol consumption behaviors in nonpatient adult populations in high-income countries.

Methods

Overview

The study protocol was registered with PROSPERO (registration number: CRD42019126074). All deviations from the protocol are explained in the Methods section. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting (see Multimedia Appendix 1 for the checklist) [15].

Data Sources

Relevant reviews were obtained through an internet-based search and a manual search. First, 4 internet-based databases (MEDLINE, Embase, PsycINFO, and Cochrane Reviews) were searched for peer-reviewed review articles published in English between January 1, 2009, and February 25, 2019. We limited our search dates, only starting in January 2009, to make our study manageable and also because we expected reviews published in the last decade to capture earlier papers. Publications were restricted to English due to the absence of translation expertise. Gray literature searches were conducted in OpenGrey, ProQuest Dissertations and Theses, Google, and targeted websites (see Multimedia Appendix 2 for search terms and the gray literature search strategy). More articles were identified by manual searches of the reference lists from excluded reviews of reviews. We did not search in study registries, as we were looking for systematic reviews. We did not conduct full hand searches or consult experts to ascertain the literature for pragmatic and logistical reasons.

XSL•FO
RenderX

Review Selection

The reviews were screened using a three-stage process. A total of 2 reviewers (NG and AY) examined titles and discarded reviews that did not meet the inclusion criteria (Textbox 1). Each reviewer then independently screened the abstracts of 10% of the remaining reviews to identify studies that potentially met the inclusion criteria. Interreviewer agreement on inclusion was also assessed. Reviewers disagreed on 11 of 41 decisions. All disagreements were resolved through discussion. AY screened all the remaining abstracts. The relevant review articles were then obtained in full and screened independently for eligibility by NG and AY. Any disagreement over eligibility was resolved through discussion with a third reviewer (TC or BR).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

1. Study type: systematic reviews (whose reporting of the evidence could be either by narrative synthesis or by meta-analysis) that reported on the effectiveness of digital interventions in changing health-related behavior and/or health outcomes. We did not restrict by study design of the included studies within the systematic reviews.
2. Population: this included adult nonclinical populations. We aimed to assess the effectiveness of digital interventions in relation to CVD prevention relevant to the NHS Health Check program, which is offered to adults aged between 40 and 74 years (England, United Kingdom). Where populations were mixed, we included the review if the population of interest could be isolated. Where impossible to isolate the population of interest, reviews were included if ≥50% of the studies were of relevant populations.
3. Intervention: we included digital interventions targeting behaviors related to diet, physical activity, smoking, and/or alcohol consumption. Digital interventions include interventions delivered over the internet (web-based or websites), mobile telephone interventions (including texts and mobile apps), social media, computer-delivered interventions, and wearable technology. Interventions incorporating both digital and face-to-face components were also included.
4. Comparator: there were no restrictions. We extracted information about the comparators where available, allowing us to review effectiveness compared with both nondigital interventions and nonintervention controls.
5. Outcome: this included behavioral or health outcomes related to diet, physical activity, alcohol consumption, and smoking. Reviews that considered these areas of behavior, either individually or in combination, were included.
6. Time frame for follow-up: any time frame.

Exclusion criteria

1. Study type: reviews of reviews, conference abstracts, protocols, opinion pieces, and commentaries. We excluded reviews of reviews because we expected that most of the reviews gathered in a review of reviews would already be included in our study. Therefore, including reviews of reviews would have led to double counting of some information.
2. Population: reviews that only considered any of the following in ≥50% of included studies: children and adolescents, students, adults aged <40 years, pregnant women, management of existing CVD or other health conditions, and low- to middle-income countries. These criteria were selected to protect the ecological validity of this review, as relevant to the NHS Health Check.
3. Intervention: reviews of nondigital interventions. We did not consider television, radio, or telephone calls to be digital, as they are not often used in digital interventions for public health.
4. Comparator: no exclusion criteria.
5. Outcome: feasibility, acceptability, participation, and engagement only.
6. Superseded: this included reviews updated by subsequent reviews that included all the same studies as the original.

Study Quality Assessment

Review confidence was critically appraised independently by 2 reviewers (NG and AY) for a 10% subsample of the included publications, using the AMSTAR-2 tool [16]. The findings were discussed to check for consistency. The remaining articles were divided and assessed by NG or AY. Any uncertainty was resolved through discussion with a third reviewer (BR).

Data Extraction

Data were extracted using a standardized form (Multimedia Appendix 3). We extracted data on the following predefined review components: objective, population, inclusion and exclusion criteria, search date, included studies (number, type, and countries), follow-up, method of synthesis, results and findings, and comparator. A total of 2 reviewers (NG and AY) extracted the data independently for 10% of the publications. Discrepancies were resolved through discussion. The remaining publications were divided among both reviewers. During data extraction, we also noted information about the control condition and any information on a comparison of effectiveness of no-intervention versus active controls. Data on adherence and attrition were also recorded, where available. The data extraction form has been presented in Multimedia Appendix 3.

Analysis

We conducted a systematic narrative synthesis using extracted data from included articles. No statistical analyses were conducted and meta-analysis was not possible with the included articles. We have presented results in the following categories: diet, physical activity, diet and physical activity combined, smoking, alcohol consumption, and multiple areas of behavior (all combinations other than diet and physical activity). Where there were enough reviews, we grouped by mode of delivery,

XSL•FO
RenderX

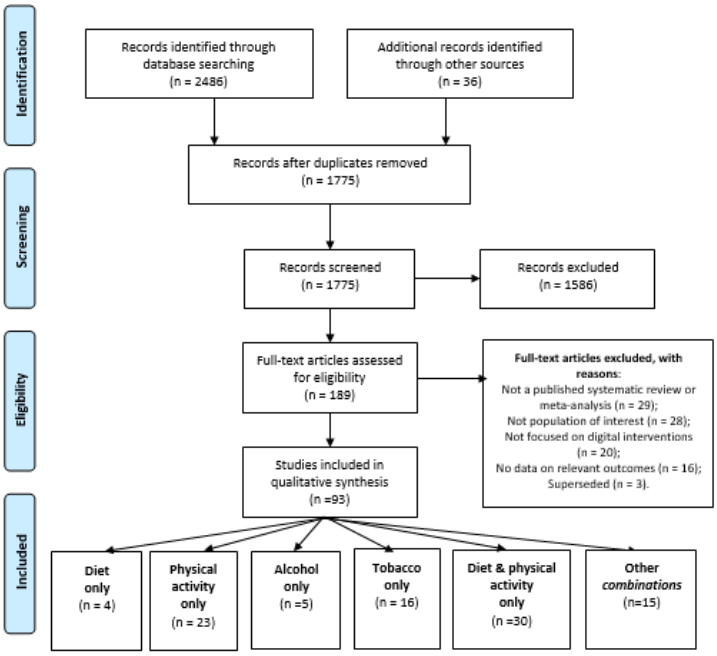
especially internet (including email and interventions that require accessing a website) and mobile phone (including apps and SMS text messaging interventions); social media was categorized separately from internet, mainly because there are enough papers to make the subdivision worthwhile but also because the social aspect may differentiate social media interventions from other forms of internet interventions, so that it is appropriately considered a subclass [17]. When reporting effect sizes, for Cohen *d*, Hedges *g*, and other measures of standardized mean difference (SMD), we followed the convention that 0.2 is a small effect size, 0.5 is a medium effect size, and 0.8 is a large effect size [18]. For risk ratios (RRs), we classified 1.22 as small, 1.86 as medium, and 3.00 as large; for odds ratios (ORs), these were 1.32, 2.38, and 4.70, respectively [19].

Results

Overview

Searches identified 1739 potentially relevant records. After screening the titles and abstracts, 154 articles were retrieved in full. An additional 36 articles were identified through hand searches and gray literature searches. In total, 94 reviews met the inclusion criteria (Figure 1). A list of reviews excluded after full-text screening is provided in Multimedia Appendix 4. We were unable to retrieve one gray literature item by May 13, 2019, and it was therefore excluded.

Figure 1. PRISMA flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Review Characteristics

The included reviews examined the effectiveness of digital interventions on diet only [20-23], physical activity only [24-46], diet and physical activity combined [47-76] (for some it was possible to extract separate results about diet and physical activity behaviors, whereas some reported more general results on weight loss outcomes), alcohol consumption [77-82], and smoking cessation [83-98]. A further 15 reviews examined the effectiveness of digital interventions on a combination of our 4 target areas of behavior (not diet and physical activity; again, sometimes it was possible to extract separate information for each target area but other times, the results were only reported in combination) [99-113]. Some reviews covered a number of areas of behavior because their research questions focused on a health outcome (eg, CVD) or a mode of intervention delivery (eg, internet interventions) rather than a behavior. Where extracting information on an area of behavior from a combination review was possible, we included the relevant data in the results for that area. For a breakdown of the reviews, with the number found in each area, for each mode of intervention and type of control see Multimedia Appendix 5.

The populations reviewed were general (nonclinical) adult populations. However, the diet and physical activity reviews were often restricted to populations of individuals with overweight or obesity. The alcohol reviews were often restricted to *problem drinkers*, defined with reference to local guidelines [77,79,80], questionnaire scores [81], or reduced productivity at work [78]. There was a range of modes of delivery, including mobile app, SMS text messaging, social media, pedometer, wearable, and interactive computer program. The reviews included both active and nonactive or minimal intervention controls, many pooling both types, but where possible, we tried to extract separate information about effectiveness of active compared with nonactive controls. We regarded the provision of educational materials as a nonactive or minimal control. More reviews included behavioral outcomes (such as fat consumption, fruit and vegetable consumption, physical activity, alcohol consumption, smoking cessation, and smoking abstinence) than health outcomes (such as weight loss, BMI, and waist

XSL•FO
RenderX

circumference). Multimedia Appendix 6 provides the key characteristics of all included studies.

Review Confidence Ratings

The confidence rating of each review is presented in the study characteristics table in Multimedia Appendix 6, and a summary of confidence ratings is provided in Table 1; 84% (79/93) of reviews were rated as critically low. During the completion of the AMSTAR-2 tool, reviewers noted that most reviews failed to satisfy items 4 (including and justifying a publication language inclusion criterion) and 7 (providing a list of excluded reviews and justifications). Both items are considered critical for systematic reviews but not for meta-analyses. A modified rating was produced alongside the original AMSTAR-2 rating, which did not classify either of these flaws as critical, to see whether a variation among reviews would be revealed. However, little change was observed, with only 5 reviews moving from a rating of critically low to low.

Table 1. Risk of bias: a summary of AMSTAR-2 confidence ratings.

Category	Confidence rating (modified rating)			
	High	Moderate	Low	Critically low
Diet	0 (0)	0 (0)	1 (1)	3 (3)
Physical activity	0 (0)	1 (1)	0 (1)	20 (19)
Diet and physical activity	0 (0)	1 (1)	1 (3)	28 (26)
Alcohol	0 (0)	0 (0)	1 (1)	4 (4)
Smoking	0 (0)	2 (2)	4 (4)	10 (10)
Other	0 (0)	1 (1)	2 (4)	14 (12)
Total	0 (0)	5 (5)	9 (14)	79 (74)

Effectiveness of Interventions

Table 2 summarizes the effectiveness of the different types of interventions.

XSL•FO
RenderX

Table 2. Summary of the effectiveness of different types of interventions.

Findings	Behavioral categories				
	Diet	Physical activity	Diet and physical Activity	Smoking	Alcohol	Other combinations
Number of reviews	• 20 reviews, of which 4 were diet-only	• 45 reviews, of which 22 were physical activity-only	• 35 reviews, of which 29 were on weight loss, specifically combining diet and physical activity, and 3 where results on weight loss were extractable from other combinations	• 28 reviews, of which 16 were smoking-only	13 reviews, of which 6 were alcohol-only	11
Effectiveness compared with mixed (active and nonactive) controls	• Small effect of internet interventions on pooled behaviors but heterogenous outcome measures and effect sizes • Some small effects of mobile phone interventions on behavior, but no effects on weight loss • Mixed effects of social media but no effects on weight loss • Mixed effects of computer-delivered interventions, and positive effects were not clinically significant	• Mixed evidence for the effect of internet interventions on physical activity; where there were effects, they were small and there was heterogeneity between studies • Mixed effects of social media, with meta-analyses not finding significant effects • Better evidence for effects of mobile phone interventions, with most studies in narrative reviews showing effectiveness, with effect sizes for SMS text messaging interventions ranging from small to medium • Mixed evidence for the effect of exergaming • Mixed evidence on computer-delivered interventions (2 reviews); the meta-analysis found small effects • There was favorable evidence from reviews assessing a variety or a combination of interventions	• Positive effects of internet interventions for weight loss and BMI, especially when part of blended interventions • Mixed effects of social media interventions • Medium-sized effects of mobile interventions on weight loss • Mixed results from reviews covering a variety of interventions, around half of the interventions were found to be effective	• Small positive effects for internet, mobile, and computer-delivered interventions. Mixed evidence about interventions from reviews that covered multiple types of digital technologies	Small and medium positive effects	Small positive effects on behavior and on health outcomes for internet, computer-delivered, SMS text messaging, and prompts delivered by all methods; however, results were not favorable for apps

XSL•FO
RenderX

Findings	Behavioral categories					
	Diet	Physical activity	Diet and physical Activity	Smoking	Alcohol	Other combinations
Comparison to no or minimal intervention	Little evidence available and it was mixed: • One meta-analysis of voice-response interventions found no effect • One meta-analysis of computer-tailored interventions found a small effect	Mixed results: • Mixed results for internet and mobile interventions • One review of social media interventions, finding increased in steps taken but not energy expenditure, total physical activity, or moderate-to-vigorous physical activity • Small positive effect of interactive voice response and computer-delivered interventions	• More effective than minimal controls for internet and computer-based interventions, but effect sizes were small	• Small positive effects for internet-based interventions and for digital interventions not broken down by mode of delivery), mobile interventions more effective than completely passive controls and as effective as minimal controls	Effective compared with no-intervention controls	Effective compared with no-intervention controls, with small effects for internet and medium effects for SMS text messaging interventions
Comparison with active controls	• Mixed; in narrative syntheses (of internet, mobile, and combinations of interventions), less than 50% of studies in each review found significant improvements	• Increase in physical activity for internet-based interventions • Mixed results for mobile interventions • Promising results about wearables	• Internet-based interventions not more effective than active controls, unless as a part of blended interventions • Computer-delivered interventions led to less weight loss than in-person treatment	• No evidence that digital interventions are any more effective than active controls	No evidence that digital interventions are any more effective than active controls; mixed evidence about whether active controls are more effective than digital	Very small effects for internet and small effects for computer-delivered and SMS text messaging interventions
Sustainability	• Relatively few studies with follow-ups • Digital interventions were largely effective over a 3- to 6-month period • Evidence was more mixed in the long term (>12 months)	• Relatively few studies with follow-ups • Some evidence for sustainability of internet and combinations of digital interventions at 6-month follow-up • Some evidence that the effect of mobile interventions is short lived	• Few studies, but those found that effectiveness declined over time		Mixed results about whether the effect was sustained. No follow-ups exceeded 12 months. Evidence suggests effects have diminished by this point	Mixed evidence on when the effect size peaks (short or medium term); effects decrease but still exist at 12 months

XSL•FO
RenderX

Findings	Behavioral categories					
	Diet	Physical activity	Diet and physical Activity	Smoking	Alcohol	Other combinations
				Mixed evidence: • Effects of internet-based interventions sustained up to 6 months, but there is disagreement about whether effects are maintained at the 12-month follow-up • Mixed evidence on the sustainability of mobile interventions • No evidence of sustainability of computer-delivered interventions • Reviews that surveyed a range of digital interventions without differentiating by mode of delivery found sustained effects, even at the 18-month follow-up		

Diet

Review Characteristics

A total of 20 reviews reported findings on diet behaviors or a weight loss outcome that resulted from an intervention that only targeted diet. The breakdown of their characteristics is shown in Table S1.

Effectiveness of Digital Interventions on Diet Compared With Mixed (Active and Nonactive) Controls for Diet Behaviors

The effectiveness of digital interventions in improving results related to diet was at best mixed, for both behavioral and health outcomes (Table 3); where improvements were reported, the effect sizes were typically small. This was the case across all modes of delivery.

XSL•FO
RenderX

Table 3. Results of reviews on diet, ordered by type of control and mode of delivery of intervention.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Dietary outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed (active and nonactive) controls							
Afshin et al (2016) [99]	20 (224)	Narrative synthesis of RCTs[a] and quasiexperimental studies	Internet and mobile	Various dietary behaviors, fruit and veg intake	1 week-2 years	70% (14/20) studies found significant dietary improvements. Effect sizes varied due to heterogeneity in dietary targets. The intake of fruit increased by approximately 1 serving/day. Two RCTs assessed mobile - based interventions and fruit or vegetable intake; each found significant improvement (by 2 and 4 servings/day)	Critically low
Aneni et al (2014) [100]	9 (29)	Narrative synthesis of RCTs	Internet	Dietary intake	6-24 months	44% (4/9) high-quality studies demonstrated improvements in diet	Critically low
Hou et al (2013) [108]	7 (38)	Narrative synthesis of studies with comparison or control groups	Internet	Dietary intake and fruit and vegetable intake	Not reported	All the studies examining nutrition alone and nutrition and other factors in addition to PA[b] reported increases in healthy dietary behaviors in the intervention groups. Interventions showed promising effects at 12 months for reductions in body fat, weight, and dietary fat intake	Critically low
Lustria et al (2013) [110]	10 (40)	Meta-analysis of experimental and quasiexperimental studies	Internet-tailored interventions	Fruit and vegetable intake and saturated fat intake	6 weeks-6 months	Small effect sizes, Cohen d=0.223 (k=4; 95% CI 0.11 to 0.33; P<.001)	Critically low
Maon et al (2012) [63]	8 (26)	Narrative synthesis of RCTs	Internet	Fruit and vegetable consumption	6 weeks-2 years	50% (4/8) studies investigating healthy eating habits reported positive changes such as increased fruit or vegetable consumption	Critically low
Webb et al (2010) [113]	10 (85)	Meta-analysis of RCTs	Internet	Dietary behavior	Not reported	Small effect sizes on behavior were observed for interventions that targeted only dietary behavior (Cohen d_+=0.20; k=10; 95% CI 0.02 to 0.37)	Critically low
DiFillipo et al (2015) [21]	3 (3)	Narrative synthesis of RCTs	Mobile	Weight loss	8 weeks-6 months	100% (3/3) of studies found a numerical tendency to weight loss compared with the control; only 33% (1/3) of studies was statistically significant; this study showed an increase in self-monitoring	Critically low
McCarroll et al (2017) [23] (diet only)	21 (23)	Narrative synthesis of RCTs	Mobile	Healthy eating and weight loss (diet only)	1-24 months	Small positive effects of interventions on healthy eating (5/8, 63% of trials) and weight loss (5/13, 38% of trials), but studies were judged to be of poor quality	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Dietary outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Lyzwinksi et al (2014) [60]	5 (14)	Meta-analysis of RCTs	Mobile	Fruit and veg intake, energy density	8 weeks-12 months	100% (3/3) of studies found an improvement in fruit and vegetable intake; 100% (2/2) studies found an improvement in energy density and eating behavior	Critically low
Palmer et al (2018) [111]	3 (71)	Meta-analysis of RCTs	Mobile	Saturated fat intake, BMI, salt intake, weight from diet-only interventions	Not reported	2 (100%) studies showed no effect of interventions on weight or dietary intake	Moderate
Elaheebocus et al (2018) [56]	20 (134)	Narrative synthesis of RCTs	Social media	Body weight	6-48 months	75% (15/20) of studies using web-based social networks had positive results for dietary outcomes	Critically low
Mita et al (2016) [65]	12 (16)	Meta-analysis of RCTs	Social media	Body weight and fruit and vegetable intake	Not reported	No significant differences (SMD[c] −0.14; 95% CI −0.28 to 0.01), with similar findings for body weight (SMD 0.07; 95% CI −0.17 to 0.20) and fruit and vegetable intake (SMD 0.39; 95% CI −0.11 to 0.89)	Moderate
Williams et al (2012) [75]	5 (22)	Meta-analysis of RCTs	Social media	Body weight and dietary fat	3-24 months	No significant differences in changes in weight (SMD 0; 95% CI −0.19 to 0.19; 10 studies) however, pooled results from 5 studies showed a significant decrease in dietary fat consumption with social media (SMD −0.35; 95% CI −0.68 to −0.02)	Critically low
Wieland et al (2012) [74]	18 (18)	Meta-analysis of RCTs, quasi-RCTs, cluster RCTs, and quasi-experimental studies	Interactive, computer-based	Fat intake, calorie intake, total fiber, fruit and vegetable intake; weight, BMI, and waist circumference	4 weeks-30 months	Computer-based interventions led to greater weight loss than minimal interventions (MD[d] −1.5 kg; 95% CI −2.1 to −0.9) but less weight loss than in-person treatment (MD 2.1 kg; 95% CI 0.8 to 3.4; one trial); in the 3- to 6-month follow-up, there was a significant decrease in percentage calories from fat (MD −1.1%), and improved fiber intake (dietary fiber score MD 1.3); there was no significant effect on energy intake at 30 months	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Dietary outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Harris et al (2011) [22] (diet only)	25 (43)	Meta-analysis of RCTs	Interactive, computer based	Fat intake, fruit and vegetable consumption, fiber intake, and energy intake	0-12 months	Interventions did not produce clinically significant changes in dietary behavior: fruit and vegetable intake had a weighted MD of 0.24 servings per day (95% CI 0.04 to 0.44 servings; *P*=.02), total energy consumed from fat had a weighted MD of –1.4% (95% CI –2.5% to –0.3%; *P*=.01); there were no significant weighted MDs in intake of total fat, saturated fat, daily dietary fiber, or daily energy; there were no significant effects on fruit and vegetable consumption and BMI at 12 months	Low
Carvalho de Menzes et al (2016) [53]	15 (18)	Narrative synthesis of all designs	Various	Fat consumption and fruit and vegetable consumption	1-36 months	Among the changes in eating habits, statistically significant reductions were observed in the consumption of fat (total fat, saturated fat, and transfat), and there was an increase in the ingestion of fruit and vegetables	Critically low
Nonactive controls							
Krebs et al (2010) [109]	51 (76)	Meta-analysis of RCTs	Computer-tailored	Dietary intake of fat, vegetables, and fruits	1-24 months	The mean effect size for dietary fat reduction was *g*=0.22 (95% CI 0.18 to 0.26); mean effect size for fruit and vegetable intake was *g*=0.16 (95% CI 0.10 to 0.21).	Low
Tsoli et al (2018) [73]	2 (15)	Meta-analysis of RCTs	Interactive voice response	Diet	6 weeks-12 months	No statistically significant effect on behaviors related to diet (Hedges *g*=0.130; 95% CI −0.088 to 0.347; k=2; *P*=.24)	Critically low
Active controls							
Burke et al (2011) [20] (diet only)	5 (24)	Narrative synthesis of RCTs and descriptive studies	Personal digital assistants	Weight loss (diet only)	3-24 months	In all studies, self-monitoring was significantly associated with weight loss, but there was minimal evidence to support the use of personal digital assistants over other methods of self-monitoring; they may result in long-term reduced risk of weight regain after 18 months	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Dietary outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Covolo et al (2017) [104]	18 (40)	Narrative synthesis of RCTs	Mobile	BMI and waist circumference; fruit and vegetable intake, and high-sugar food intake	6 months-2 years	56% (10/18) of RCTs found no difference between intervention and control group; 33% (6/18) studies showed a significant increase in the consumption of fruit and vegetables and reduced sugar-sweetened beverage consumption	Critically low

[a]RCT: randomized controlled trial.

[b]PA: physical activity.

[c]SMD: standardized mean difference.

[d]MD: mean difference.

Of the 6 reviews covering internet interventions, 2 meta-analyses found small, favorable effects of internet interventions on dietary behavior, Cohen *d*=0.223 [110] and Cohen *d*+=0.20 [113]. A total of 4 narrative syntheses found mixed results with heterogeneous target outcomes [63,99,100,108]. The most common dietary target across the studies was fruit intake, which increased by approximately 1 serving per day [99].

Across 6 reviews of mobile interventions, there was some evidence of positive effects on dietary behaviors, especially fruit and vegetable intake, but no statistically significant effects on weight loss [21,23,60,99,104,111]. One of these reviews [99] reported a previous systematic review (not covered in this paper) that found no significant change in calorie intake or the consumption of sugar-sweetened beverages in 4 trials evaluating diet.

A total of 3 social media reviews had mixed findings on dietary outcomes, such as fruit and vegetable intake and fat intake, but the 2 reviews with meta-analyses found no differences in weight [65,75].

The effects of interactive computer interventions were equivocal and not clinically significant [22,74].

Effect of Digital Interventions on Diet Compared With Nonactive Controls

There was little evidence about the effectiveness of digital interventions on diet compared with no intervention or minimal intervention controls, and the available evidence was mixed (Table 3). One meta-analysis found small effects on fruit and vegetable intake (g=0.16) and dietary fat reduction (g=0.22) [109], but another meta-analysis found no statistically significant effects on dietary behaviors [73].

Effect of Digital Interventions on Diet Compared With Active Controls

The relative effectiveness of digital interventions on diet compared with active comparators was mixed, with approximately half of the studies or less in narrative syntheses, showing that digital interventions were effective compared with active controls (Table 3). This was true across all modes of delivery: internet [99], mobile [99,104], and combined interventions [99].

Sustainability of Effects on Diet at Follow-Up

Reviews included studies that ranged from single-contact interventions to 5-year follow-ups (Table 3). Relatively few reviews reported on follow-ups; of those that did, about half reported follow-ups in the medium term (3-6 months) and half in the long term (≥12 months).

Where reported, digital interventions were generally found to be effective for over 3 to 6 months. Several reviews have found positive results at 6 months [20,21,74]. However, this finding is not universal [111].

Long-term findings were more mixed in the 5 reviews that investigated them. Two reviews suggested promising effects at 12 months [108] and 18 months [20]. However, 3 reviews found no significant effects at 12 months [22], 24 months [111], and 30 months [74].

Physical Activity

Review Characteristics

We included the findings on physical activity–related outcomes from 45 systematic reviews. The breakdown of their characteristics is shown in Multimedia Appendix 6.

Effectiveness of Digital Interventions on Physical Activity Compared With Mixed (Active and Nonactive) Controls

The effectiveness of digital interventions was mixed across all modes of delivery, apart from mobiles, for which the evidence was consistently positive (Table 4).

XSL•FO
RenderX

Table 4. Results of reviews on physical activity, ordered by type of control and further ordered by mode of delivery of intervention.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed controls (active and nonactive)							
Aalbers et al (2011) [47]	2 (10)	Narrative synthesis of randomized and nonrandomized pre-post controlled trials	Internet	Total PA[a] and MV-PA[b]	1.5-6.5 months	2 studies reported the effects of digital interventions, one was less effective on MVPA than a nonactive control and the other demonstrated a small positive effect on total PA[b] (*P*=.001).	Critically low
Aneni et al (2014) [100]	9 (29)	Narrative synthesis of RCTs[c]	Internet	PA measures	6-24 months	No improvement was seen in virtually all the studies with PA outcome, only 11% (1/9) of studies demonstrated a significant intervention effect on PA.	Critically low
Bottorf et al (2014) [26] (PA only)	8 (35)	Narrative synthesis of all study types	Internet	Changes in PA; step count; self-reported walking; BMI; waist circumference; weight	3-12 months	63% (5/8) of studies demonstrated that PA significantly increased in the internet-based interventions, 2 studies showed a nonsignificant difference, and one showed that the effects were indeterminable.	Critically low
Davies et al (2012) [29] (PA only)	34 (34)	Meta-analysis of experimental design studies	Internet	PA	2-52 weeks	The estimated overall mean effect of internet-delivered interventions on PA was Cohen *d*=0.14 (*P*<.001). Homogeneity tests from the fixed-effect analysis revealed significant heterogeneity across studies (*Q*=73.75; *P*<.001). The overall mean effect for sustained PA at least 6 months postintervention (n=11) resulted in a small but significant effect size Cohen *d*=0.11 (*P*<.01).	Critically low
George et al (2012) [33] (PA only)	2 (14)	Narrative synthesis of all study types	Internet	Step count; health status; BMI; weekly PA	2-8 months	Increase in PA in 100% (2/2) of online interventions where participants were in competitive teams, including one that showed an increase in step count. Poor quality evidence.	Critically low
Hou et al (2013) [108]	7 (38)	Narrative synthesis of trials with comparison or control group	Internet	Level of physical activity	0-12 months	86% (6/7) of interventions were successful in the studies focusing primarily on PA.	Critically low
Jahangiry et al (2017) [35] (PA only)	21 (22)	Meta-analysis of controlled trials	Internet	MVPA; walking; step count (pedometer)	1-20 weeks	36% (5/14) of rials reporting MVPA, 50% (3/6) of trials reporting step count, and 29% (4/14) of studies reporting minutes walking showed significant increases. The interventions were influenced by the age of participants and trial length.	Critically low
Lustria et al (2013) [110]	12 (40)	Meta-analysis of experimental and quasiexperimental studies	Internet	Levels of PA	4 weeks-24 months	The sample size–weighted mean effect size for studies on PA was not significant Cohen *d*=0.059 (k=12; 95% CI −0.02 to 0.14).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Maon et al (2012) [63]	13 (26)	Meta-analysis and narrative synthesis of RCTs	Internet	PA levels, sedentary behavior, and MV-PA	6 weeks-2 years	54% (7/13) of studies showed statistically significant effects on PA levels, such as increased walking or decreased sedentary behavior. However, a meta-analysis on 4 studies with extractable data for the outcome of moderate-to-vigorous weekly PA found a not statistically significant improvement: SMD[d] 0.15 (95% CI 20.06 to 0.35; *P*=.16) Duration of studies and effects: 10% (3/30) of studies showed positive effects when outcomes were measured immediately after the end of the interventions. In total, 37% (11/30) of studies that lasted 3 months or less demonstrated positive outcomes; 43% (13/30) of studies with an intervention of 3-6 months showed positive results; and only 10% (3/30) interventions that lasted longer than 6 months were reported to have positive results.	Critically low
Webb et al (2010) [113]	20 (85)	Meta-analysis of RCTs	Internet	Level of PA	3-12 months	Small effects on behavior were observed for interventions that targeted only PA (Cohen *d*+=0.24; k=20; 95% CI 0.09 to 0.38).	Critically low
Buchholz et al (2013) [27] (PA only)	10 (10)	Narrative synthesis of RCTs, quasiexperimental and, single groups	Mobile (SMS text messaging)	Self-reported frequency or pedometer-reported steps and level of PA	3-52 weeks	Effect sizes across all studies were positive; the median effect size was 0.5 (medium) but heterogeneous. Sample sizes were small.	Critically low
Elavsky (2018) [31] (PA only)	50 (52)	Narrative synthesis of RCTs and pre-post studies	Mobile	PA and sedentary behavior	<3 months	59% (17/29) of RCTs and 62% (13/21) of pre-post studies supported the effectiveness of mobile interventions to improve PA, and 9 (5 of 10 RCTs and all 4 pre-post) of 14 (64%) studies reduced sedentary behavior.	Critically low
Lyzwinksi et al (2014) [60]	9 (14)	Meta-analysis of RCTs	Mobile	Levels of PA	8 weeks-12 months	Trials mostly found that PA levels increased in the intervention groups relative to the control groups.	Critically low
Maher et al (2014) [61]	4 (10)	Narrative synthesis of studies with comparator group (control or within subject)	Mobile	Levels of PA	8 weeks-24 months	25% (1/4) of studies demonstrated a significant change in PA Cohen *d*=0.84 (95% CI −0.49 to 1.19).	Critically low
Muntaner et al (2015) [41] (PA only)	11 (11)	Narrative synthesis of all study types	Mobile	PA; exercise	2-24 weeks	55% (6/11) of articles included in this review reported significant increases in PA levels.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
O'Reilly et al (2013) [42] (PA only)	12 (22)	Narrative synthesis of RCTs	Mobile	PA; sedentary behavior; BMI; blood lipids; blood pressure; QoL[e]; adverse effects	Not reported	75% (9/12) of studies reported significant changes in PA or sedentary behavior.	Critically low
Palmer et al (2018) [111]	15 (71)	Meta-analysis of RCTs	Mobile	Level of PA	3 months	Trials of PA interventions reporting outcomes at 3 months showed no benefits.	Moderate
Schoeppe et al (2016) [68]	10 (27)	Narrative synthesis of RCTs, randomized trials, controlled trials, and pre- and poststudies	Mobile	PA; sedentary behavior	1-24 weeks	59% (13/22) of studies reported significant improvements in levels of PA; 20% (1/5) of studies reported a significant change in sedentary behavior.	Critically low
Elaheebocus et al (2018) [56]	25 (134)	Narrative synthesis of RCTs	Social media	Body weight	6-48 months	76% (19/25) of studies using online social networks had positive results.	Critically low
Mita et al (2016) [65]	11 (16)	Meta-analysis of RCTs	Social media	PA; weight change	1-12 months	For PA, significant mean difference 0.07; 95% CI −0.25 to 0.38; k=11.	Moderate
Williams et al (2012) [75]	12 (22)	Meta-analysis of RCTs	Social media	PA	3-24 months	Meta-analysis showed no significant differences in changes in PA (SMD 0.13; 95% CI −0.04 to 0.30; k=12).	Critically low
Willis et al (2017) [76]	3 (5)	Narrative synthesis of all study types	Social media	Total PA	8 weeks-6 months	Only one study reported significant changes in levels of PA, when the web-based social network intervention included an online support group.	Critically low
Johnson (2017) [37]	10 (19)	Narrative synthesis of RCTs	Active gaming	Behavioral and cognitive outcomes	Not reported	Findings were largely positive for behavioral impacts, specifically the impact of gamification for PA: 80% (8/10) positive and 20% (2/10) mixed.	Critically low
Peng et al (2012) [43] (PA only)	4 (12)	Narrative synthesis of all study designs	Active gaming	Heart rate; energy expenditure; and oxygen uptake	6-12 weeks	Evidence does not support active video games as an effective tool to significantly increase PA or exercise attendance.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Street et al (2017) [46] (PA only)	9 (9)	Narrative synthesis of studies with comparison or control groups	Active gaming	PA; maximum oxygen uptake; power; blood pressure; body mass; body weight; body fat; BMI; balance; speed; and strength	6-12 weeks	Moderate-to-high exergaming participation was associated with statistically significant improvements in anthropometric outcomes but low participation was not associated with anthropometric changes. 38% (3/8) studies that investigated anthropometric outcomes, including BMI and body fat, found a statistically significant improvement, all 3 studies showed positive health outcomes associated with moderate-to-high participation in exergaming; 100% (3/3) of studies that reported on PA frequency reported higher frequency in the exergaming condition; however, a different 100% (3/3) of studies that reported on overall PA found no statistically significant increases.	Critically low
Wieland et al (2012) [74]	4 (18)	Meta-analysis of RCTs, quasi-RCTs, cluster RCTs, and quasiexperimental studies	Computer-delivered	Steps per day and minutes walked continuously	4 weeks-30 months	No studies demonstrated statistically significant effects on PA.	Critically low
Afshin et al (2016) [99]	33 (224)	Narrative synthesis of RCTs and quasi-experimental studies	Various: internet and mobile	Level of PA	1 week-5 years	88% (29/33) of studies reported significant improvement in PA; 83% (5/6) of phone interventions were effective, including 66% (2/3) of SMS text messaging interventions, 100% (2/2) of apps, and 100% (1/1) of automated voice response.	Critically low
Carvalho de Menzes et al (2016) [53]	13 (18)	Narrative synthesis of all study designs	Various: email, telephone, websites	Level of PA	1-36 months	Most studies demonstrated statistically significant improvements in the level of PA.	Critically low
Hakala et al (2017) [34] (PA only)	13 (23)	Meta-analysis of RCTs	Various: mobile, text messages, pedometers, wearables, email	PA: self-reported or using an accelerometer or pedometer	3 weeks-24 months	No differences were observed between the experimental and control groups (risk ratio 1.03; 95% CI 0.92 to 1.15; *P*=.57).	Critically low
Muellmann et al (2018) [39] (PA only)	13 (20)	Narrative synthesis of experimental designs and quasiexperimental studies	SMS text messaging and internet	PA and number of steps per day	4 weeks-24 months	75% (3/4) of studies using mobile phones demonstrated significant differences in the level of PA or steps per day (mixed controls). In 100% (9/9) of studies, internet interventions significantly increased PA compared with nonactive controls.	Critically low
Muller and Khoo (2014) [40] (PA only)	4 (16)	Narrative synthesis of RCTs and quasiexperimental studies	Various: internet and mobile	PA	1 week-18 months	75% (3/4) of studies reported significant improvements in PA; 25% (1/4) of studies reported nonsignificant decrease in PA.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Stephenson et al (2017) [45] (PA only)	15 (17)	Meta-analysis of RCTs	Various: mobile messaging, mobile apps, website, wearable technology	Sedentary behavior	5 days-24 months	Interventions using computer and mobile and wearable technologies can be effective in reducing sedentary behavior. Effectiveness appeared most prominent in the short-term and lessened over time. Meta-analysis of 88% (15/17) of RCTs suggested that computer, mobile, and wearable technology tools resulted in a mean reduction of −41.28 min/day of sitting time (95% CI −60.99 to −21.58; I^2=77%). The pooled effects showed mean reductions at short (≤3 months), medium (>3 to 6 months), and long-term (>6 months) follow-up of −42.42 min/day, −37.23 min/day, and −1.65 min/day, respectively.	Critically low
Nonactive controls							
Jenkins et al (2009) [36] (PA only)	5 (22)	Narrative synthesis of RCTs	Internet	PA	0-24 months	Results were mixed; internet interventions can be effective, compared with control conditions, although poor compliance was an issue. 50% (2/4) studies reported an increase in PA compared with nonactive controls while 2 studies found no difference.	Critically low
Bock et al (2014) [24] (PA only)	4 (50)	Narrative synthesis of RCTs and quasiexperimental studies	Internet or computer	Weekly PA, proportion of sufficiently active persons; step counts	0 weeks-3 years	Interventions had a nonsignificant, positive effect on PA (*P*=.88).	Critically low
Krebs et al (2010) [109]	25 (76)	Meta-analysis of RCTs	Computer-delivered	Minutes of PA	1-18 months	The mean effect size was *g*=0.16 (95% CI 0.10 to 0.21).	Low
Bort-Roig et al (2014) [25] (PA only)	5 (26)	Narrative synthesis of comparative and pre-postdesign	Mobile	PA (steps); energy expenditure; body weight and body fat; blood pressure and cholesterol; QoL	2 weeks-6 months	80% (4/5) of studies assessing PA intervention effects reported PA increases, with mean PA increases ranging from 800 to 1104 steps/day. Studies were small with differences in baseline characteristics.	Critically low
Direito et al (2017) [30] (PA only)	17 (21)	Meta-analysis of RCTs	Mobile	PA, MVPA, walking and sedentary behavior	1-52 weeks	Not effective for MVPA outcomes, based only on adult studies SMD 0.14 (95% CI −0.10 to 0.37). For sedentary behavior outcomes, SMD −0.21 (95% CI −0.59 to 0.18).	Critically low
Freak-Poli et al (2013) [32] (PA only)	4 (4)	Narrative synthesis of RCTs and cluster RCTs	Wearable technology	PA; sedentary behavior; BMI; blood lipids; blood pressure; QoL; adverse effects	3-8 months	Overall, there was insufficient evidence to assess the effectiveness of pedometer interventions in the workplace. 75% (3/4) of studies compared with a minimal control group, 33% (1/3) of studies observed an increase in PA under a pedometer program, but the other two did not find a significant difference.	Moderate

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
An et al (2017) [49]	21 (22)	Meta-analysis of RCTs, pre-post studies, and cohort studies	Social media	PA, sedentary behavior	3-102 weeks	Interventions increased daily number of steps taken by 1530 (95% CI 82 to 2979). However, they were not associated with energy expenditure, total PA, or MVPA.	Critically low
Tsoli et al (2018) [73]	3 (15)	Meta-analysis of RCTs	Interactive voice responses	PA	6 weeks-12 months	Interventions led to a small but statistically significant increase in PA (g=0.254; 95% CI 0.068 to 0.439; k=3; P=.007).	Critically low
Active controls							
Beishuizen et al (2016) [102]	5 (57)	Meta-analysis of RCTs	Internet	Level of PA	4 weeks-3 months	Interventions led to an increase in PA (SMD 0.25; 95% CI 0.10 to 0.39).	Low
Covolo et al (2017) [104]	23 (40)	Narrative synthesis of RCTs	Mobile apps	Daily steps, frequency, and intensity of PA	6 months-2 years	30% (7/23) of RCTs showed a significant increase in PA in the intervention group (measured in daily steps, frequency of PA, or level of intensity), 48% (11/23) of studies did not show a significant increase, and in 21% (5/23) studies, outcome measures were inconsistent in whether there was a significant difference between intervention and control.	Critically low
Mateo et al (2015) [38] (PA only)	10 (11)	Meta-analysis of controlled trials	Mobile apps	PA, MVPA, and steps	6 weeks-9 months	Compared with the control group, use of a mobile phone app was associated with significant changes in body weight and BMI of −1.04 kg (95% CI −1.75 to −0.34; I^2=41%) and −0.43 kg/m^2 (95% CI −0.74 to −0.13; I^2=50%), respectively (k=9); however, a nonsignificant difference in PA was observed between the intervention and comparison groups (SMD 0.40; 95% CI −0.07 to 0.87; I^2=93%).	Critically low
Song et al (2018) [44] (PA only)	6 (8)	Narrative synthesis of all study types	Mobile	PA (frequency and step count); BMI; blood glucose	4 weeks-6 months	Significant effects on frequency of PA in 80% (4/5) of studies (though the effect was reported to have disappeared after the 12-week follow-up), step count in 66% (2/3) of studies, BMI in 50% (2/4) of studies, and reduction in glucose in 100% (2/2) studies.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Cheatham et al (2018) [28] (PA only)	25 (25)	Narrative synthesis of controlled clinical trials	Wearable technology	PA; BMI; weight; blood pressure; Resting Energy Expenditure; body composition; cardiovascular fitness; work productivity and absenteeism; waist circumference; blood parameters	3 weeks-24 months	An activity tracker combined with a comprehensive weight loss program may provide superior short-term (≤6 months) results than a standard weight loss program in middle aged or older adults. 80% (20/25) of studies reported higher weight loss when an activity tracker was used with a weight loss intervention.	Critically low

[a]PA: physical activity.

[b]MVPA: moderate-to-vigorous physical activity.

[c]RCT: randomized controlled trial.

[d]SMD: standardized mean difference.

[e]QoL: quality of life.

Evidence for the effectiveness of internet interventions on physical activity has been mixed. In total, 5 out of 10 reviews were positive [26,29,33,108,113], including 2 meta-analyses that found small but significant effects of internet interventions: Cohen *d*=0.14 [29] and Cohen *d*+=0.24 [113]. However, there was significant heterogeneity across studies [29]. In contrast, 5 studies were not positive [35,47,63,100,110], one of which was very unfavorable, with only 1 of 9 (11%) studies demonstrating an effect of the intervention [100]. Two meta-analyses found that the effect of internet interventions on physical activity was not significant [63,110].

A total of 4 reviews of social media interventions were mixed. In total, 2 meta-analyses of social media interventions found no significant difference in changes in physical activity [65,75], and a narrative synthesis reported mixed results [76]. However, one narrative synthesis found that 76% of studies using web-based social networks had positive results for physical activity [56].

The results of the mobile interventions to improve health were more positive. In total, 8 of 10 (80%) narrative syntheses reported a majority of positive results [25,27,31,39,41,42,60,99], with one reporting an increase of 800 to 1104 steps per day [25]. One review noted that effective interventions used SMS text messaging communication or self-monitoring [42]. A review that was specifically on SMS text messaging reported that effect sizes were all greater than 0.20, and the median was 0.50, a medium effect size [27]. In contrast, 2 narrative syntheses reported that most mobile trials did not show any benefits [61,111].

There was mixed evidence of active gaming across 3 reviews. One review found that gamification has a positive impact on physical activity and found evidence that gamification can increase motivation to exercise [37]. However, another review found a positive effect on attendance but not on physical activity or BMI [46]. A third review found that active gaming did not support increases in either physical activity or attendance [43].

Computer-delivered interventions in physical activity behaviors did not have consistent results in either weight loss or weight maintenance trials [74].

In total, 4 of the 5 reviews assessing a variety of interventions found favorable results [40,45,53,99]. One of the narrative syntheses reported a wide range of values for improvement, from 1.5 to 153 extra minutes of physical activity a week and 1000 to 2600 steps per day [99]. A meta-analysis found that computer, mobile, and wearable technology led to a mean change of −41.28 minutes per day of sitting time (a reduction in sitting time) [45]. However, one meta-analysis found no difference between the experimental and control groups [34].

Effectiveness of Digital Interventions on Physical Activity Compared With Nonactive Controls

Compared with minimal controls, evidence for the effectiveness of digital interventions has been mixed.

For internet interventions, one review found favorable evidence [39], another found unfavorable evidence [24], and a third found mixed evidence [36].

The 3 reviews of mobile interventions have also provided mixed evidence. One narrative synthesis found that interventions were effective, with 4 studies (3 pre-post and 1 comparative) reporting increases of 800 to 1104 steps per day [25]. However, another study found that mobile interventions were not effective in increasing physical activity of moderate-to-vigorous intensity or in decreasing sedentary behavior [30]. Wearables were also not very effective, with only 1 of 3 (33%) studies comparing a pedometer with a minimal control showing increased physical activity [32].

Social media–based interventions increased the daily number of steps taken by 1530 steps per day [49]. However, they were not associated with energy expenditure, total physical activity, or moderate-to-vigorous physical activity.

There were small effect sizes for both computer-delivered interventions (*g*=0.16) [109] and interactive voice response–based interventions (*g*=0.254) [73].

XSL•FO
RenderX

Effectiveness of Digital Interventions on Physical Activity Compared With Active Controls

There were mixed results compared with active controls.

A meta-analysis of internet interventions found an increase in physical activity with an SMD of 0.25 compared with active controls [102].

In total, 3 reviews of mobile phones had active controls, and there were mixed results. A meta-analysis found that the use of a mobile phone app was associated with significant changes in body weight (−1.04 kg) and BMI (−0.43 kg/m^2); however, there was no significant difference in physical activity between the 2 groups [38]. A narrative synthesis app was also not favorable for assessing changes in physical activity, with less than half of the studies showing a significant increase in physical activity in the intervention group [104]. However, another narrative synthesis of general mobile interventions found that most studies had interventions that led to changes in body weight, increases in step count, and increases in frequency of physical activity [44].

There were also promising results from a review on wearables: when an activity tracker is combined with a comprehensive weight loss program, it may provide superior short-term (≤6 months) results than a standard weight loss program in middle-aged or older adults (>30 years) [28].

Sustainability of Effects on Physical Activity at Follow-Up

There was little evidence on sustainability, as many physical activity studies did not have follow-up assessment postintervention or only had follow-ups relatively soon after the intervention end point.

There is some evidence that digital interventions can have sustained effects. A meta-analysis assessing combinations of digital technologies found that the pooled effects showed mean changes (reductions) at short (≤3 months), medium (3 to 6 months), and long-term follow-up (>6 months) of −42.42 minutes per day, −37.23 minutes per day, and −1.65 minutes per day, respectively [45]. A meta-analysis of internet interventions also found a small but significant effect on physical activity for follow-ups at least six months postintervention (Cohen *d*=0.11) [29]. The sustainability of internet interventions was also supported by a narrative synthesis that found that only 12 of 35 (34%) studies had follow-up assessments, which ranged from 7 weeks to 15 months postprogram; 10 out of 12 (83%) studies demonstrated successful maintenance of physical activity and/or secondary measures indicative of positive changes in physical activity; however, follow-up durations were primarily shorter: in 9 studies, follow-up was conducted at less than 12 months [26]. However, for mobile interventions, 2 reviews found evidence that effects tended to decrease in the long term [31], with effects disappearing after as little as 12 weeks [44].

Diet and Physical Activity (Weight Management)

Review Characteristics

A total of 35 reviews reported on both diet and physical activity. The breakdown of their characteristics is shown in Multimedia Appendix 6.

Effectiveness of Digital Interventions on Diet and Physical Activity (Weight Loss) Compared With Mixed (Active and Nonactive) Controls

Overall, digital interventions were generally found to be effective, with mobile phone interventions in particular having consistently positive results (Table 5).

XSL•FO
RenderX

Table 5. Results of reviews on diet and physical activity combined, ordered by type of control and further ordered by mode of delivery of intervention.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed (active and nonactive) controls							
Aalbers et al (2011) [47]	5 (10)	Narrative synthesis of RCTs[a] and nonrandomized pre-post controlled trials	Internet	Body weight and body weight regain	1.5-6.5 months	40% (2/5) of studies reported effect sizes on body weight with small-to-medium significant effects; 1 study reported weight regain but did not reach significance.	Critically low
Aneni et al (2014) [100]	20 (29)	Narrative synthesis of RCTs	Internet	Weight, BMI, waist circumference, and body fat	6-24 months	Modest improvements were observed in more than half of the studies with weight-related outcomes; 20 studies reported on body weight: 75% (15/20) of high quality and 5 of 20 (25%) low quality); 47% (7/15) high-quality studies reported significant improvement.	Critically low
Beishuizen et al (2016) [102]	7 (57)	Meta-analysis of RCTs	Internet	Systolic blood pressure, diastolic blood pressure, HbA_{1c}[b] level, cholesterol level, weight, and level of physical activity	3-60 months	There was a significant reduction in systolic blood pressure (MD[c] –2.66 mm Hg; 95% CI –3.81 to –1.52), diastolic blood pressure (MD –1.26 mm Hg; 95% CI –1.92 to –0.60), HbA_{1c} level (MD –0.13%; 95% CI –0.22 to –0.05), LDL[d] cholesterol level (MD –2.18 mg/dL; 95% CI –3.96 to –0.41), weight (MD –1.34 kg; 95% CI –1.91 to –0.77), and an increase in physical activity (SMD[e] 0.25; 95% CI 0.10 to 0.39).	Low
Fry et al (2009) [57]	8 (19)	Narrative synthesis of all study types	Internet	Diet and physical activity	8 weeks-30 months	There were generally positive effects of prompts; there was not enough evidence to know whether the medium in which prompts were sent through affected their effectiveness but personal contact with a counsellor did enhance effectiveness.	Critically low
Hou et al (2013) [108]	7 (38)	Narrative synthesis	Internet	Body fat, weight, and dietary fat intake	0-12 months	In 71% (5/7) of studies, intervention groups lost more body fat, body weight, and dietary fat intake and maintained higher weight loss at 12 months.	Critically low
Manzoni et al (2011) [62]	26 (26)	Narrative synthesis of all study types	Internet	Weight loss and weight loss maintenance	3-24 months	Internet-based weight loss interventions enhanced by professional feedback provided through the internet are more effective for weight loss than website-only programs but less effective than telephone counselling. 93% (13/14) of studies showed a further improvement in mean weight loss (weight maintenance) after the end of the trials.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Seo et al (2015) [69]	31 (31)	Meta-analysis of RCTs	Internet	Waist circumference	4 weeks-2 years	Internet-based interventions showed a significant reduction in waist circumference (mean change –2.99 cm; 95% CI −3.68 to −2.30; I^2=93.3%) and significantly better effects on waist circumference loss (mean loss 2.38 cm; 95% CI 1.61 to 3.25; I^2=97.2%) than minimal interventions such as information-only groups; no differences with respect to waist circumference change between internet-based interventions and paper-, phone-, or person-based interventions (mean change −0.61 cm; 95% CI −2.05 to 0.83; *P*=.42; k=31).	Critically low
Sherrington et al (2016) [70]	12 (12)	Meta-analysis of RCTs	Internet	Weight loss	3-24 months	The internet-delivered weight loss interventions providing personalized feedback resulted in an MD of 2.13 kg (*P*<.001) greater weight loss in comparison with control groups receiving no personalized feedback. Heterogeneity levels showed considerable and significant heterogeneity (I^2=99%; *P*<.001) between control groups not receiving personalized feedback and the internet-delivered weight loss interventions providing personalized feedback.	Critically low
Elaheebocus et al (2018) [56]	11 (134)	Narrative synthesis of RCTs	Social media	Body weight	6-48 months	82% (9/11) of studies using web-based social networks had positive results for weight loss.	Critically low
Maher et al (2014) [61]	5 (10)	Narrative synthesis	Social media	Weight	8 weeks-24 months	Findings were mixed, from negligible to large effect sizes for weight loss.	Critically low
Mita et al (2016) [65]	10 (16)	Meta-analysis of RCTs	Social media	Weight change	1-12 months	Meta-analysis of all trials showed no significant differences for body weight (significant mean difference 0.07; 95% CI −0.17 to 0.20).	Moderate
Williams et al (2012) [75]	10 (22)	Meta-analysis of RCTs	Social media	BMI; body weight; diet	3 months-24 months	Meta-analysis showed no significant differences in changes in weight (SMD 0; 95% CI −0.19 to 0.19; 10 studies); however, pooled results from 5 studies showed a significant decrease in dietary fat consumption with social media (SMD −0.35; 95% CI −0.68 to −0.02).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Willis et al (2017) [76]	5 (5)	Narrative synthesis of all study types	Social media	Body weight; body composition; blood pressure; and blood markers	8 weeks-6 months	100% (5/5) of studies reported a reduction in baseline weight. 60% (3/5) of studies reported significant decreases in body weight when online social networks was paired with health educator support. Only one study reported a clinically significant weight loss of 55%.	Critically low
Bacigalupo et al (2013) [50]	5 (7)	Narrative analysis of RCTs	Mobile	Weight loss and BMI	9-52 weeks	Strong evidence for weight loss in the short term with moderate evidence for the medium term.	Low
Covolo et al (2017) [104]	21 (40)	Narrative synthesis of RCTs	Mobile	BMI and waist circumference	6-12 months	62% (13/21) of studies did not find a statistical difference in changes in weight. 24% (5/21) of studies found that a mobile app was more effective compared with controls (P<.05). In 3 studies, this did not differ significantly between the 2 groups.	Critically low
Head et al (2013) [107]	3 (19)	Meta-analysis of RCTs	Mobile (SMS text messaging)	Weight	Mean 81.26 days	The weighted mean effect size for weight loss was Cohen d=0.255 (95% CI .056 to .455; P=.01; k=3).	Critically low
Liu et al (2015) [59]	9 (14)	Meta-analysis of RCTs	Mobile	Weight and BMI	3-30 months	Compared with the control group, mobile phone intervention was associated with significant changes in body weight and body mass index (weight [kg]/height (m^2) of −1.44 kg (95% CI −2.12 to −0.76) and −0.24 units (95% CI −0.40 to −0.08), respectively; no differences between shorter and longer trials (< or ≥6 months; k=22).	Critically low
Lyzwinksi et al (2014) [60]	8 (17)	Meta-analysis of RCTs	Mobile	Body weight and BMI	8 weeks-12 months	75% (6/8) of studies of mobile phone interventions found significant changes in weight favoring the mobile phone intervention groups over the controls; the meta-analysis generated a medium, significant effect size of 0.430 (95% CI 0.252 to 0.609; P≤.01), favoring mobile interventions.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Palmer et al (2018) [111]	3 (71)	Meta-analysis of RCTs	Mobile	Body weight and triglyceride levels	24 hours-6 months	There were, at best, modest benefits of diet and physical activity interventions. The effect of SMS text messaging–based diet and physical activity interventions on incidence of diabetes was pooled (risk ratio 0.67; 95% CI 0.49 to 0.90; I^2=0.0%); end point weight was pooled (MD −0.99 kg; 95% CI −3.63 to 1.64; I^2=29.4%); percentage change in weight was pooled (MD −3.1; 95% CI −4.86 to −1.3; I^2=0.3%); and triglyceride levels was pooled (MD −0.19 mmol/L; 95% CI −0.29 to −0.08; I^2=0%).	Moderate
Schoeppe et al (2016) [68]	10 (27)	Narrative synthesis of RCTs, randomized trials, controlled trials, and pre- and poststudies	Mobile	Physical activity; diet; weight status; BMI; blood pressure; sedentary behavior; and fitness	1-24 weeks	40% (4/10) of studies that measured weight reported significant improvement in weight status; apps were more successful when used alongside other intervention components than when used alone.	Critically low
Siopis et al (2015) [71]	6 (14)	Meta-analysis of RCTs, quasi-RCTs, and pre-post studies	Mobile	Body weight and BMI	8 weeks-12 months	The weighted mean change in body weight in intervention participants was −2.56 kg (95% CI −3.46 to −1.65) and in controls, −0.37 kg (95% CI −1.22 to 0.48).	Critically low
Wieland et al (2012) [74]	18 (18)	Meta-analysis of RCTs, quasi-RCTs, and quasiexperimental studies	Computer based	Weight	4 weeks-30 months	At 6 months, computer-based interventions led to greater weight loss than minimal interventions (MD −1.5 kg; 95% CI −2.1 to −0.9; 2 trials) but less weight loss than in-person treatment (MD 2.1 kg; 95% CI 0.8 to 3.4; 1 trial). At 6 months, computer-based interventions were superior to a minimal control intervention in limiting weight regain (MD −0.7 kg; 95% CI −1.2 to −0.2; 2 trials) but not superior to infrequent in-person treatment (MD 0.5 kg; 95% −0.5 to 1.6; 2 trials).	Critically low
Afshin et al (2016) [99]	35 (224)	Narrative synthesis and meta-synthesis for RCTs and quasiexperimental studies	Various: internet and mobile	Weight	3-30 months	69% (24/35) of studies reported significant improvements in adiposity following the intervention. 81% (13/16) of RCTs reported significant reductions in adiposity; using the internet in the weight loss program resulted in 0.68 kg (95% CI 0.08 to 1.29 kg) additional weight reduction over a period of 3 to 30 months; in studies finding significant weight reduction, the magnitude of weight change ranged from 1 to 6 kg after 6 months of follow-up.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Allen et al (2014) [48]	38 (39)	Narrative synthesis of randomized trials	Various: internet, messaging, chat rooms, and mobile	Weight loss	5 weeks-24 months	53% (21/39) of RCTs reported statistically significant weight loss in the intervention group as compared with the control group; the proportion varied by mode of delivery, the highest proportion of successful trials involving SMS text messaging or email (67%), followed by online chat rooms (50%), web-based (48%), and self-monitoring with technology (43%).	Critically low
Bassi et al (2014) [51]	8 (28)	Narrative analysis of RCTs	Various: internet and mobile	BMI and weight	12 months	Results were mixed; 2 studies reported significant improvements with weight loss; however, effects were typically short lived, and more weight is regained in a primarily technology-based approach, as compared with personal contact.	Critically low
Carvalho de Menzes et al (2016) [53]	18 (18)	Narrative synthesis of all study types	Various: email, telephone, face-to-face, and websites	Fat consumption, fruit and vegetable consumption, and physical activity	1-36 months	Approximately half the studies showed weight loss in the intervention group.	Critically low
Coons et al (2012) [54]	13 (13)	Narrative synthesis of RCTs	Various: PDA[f], web-based, and wearables	Body mass; BMI; BP[g]; waist circumference; RHR[h]; physical activity; body fat percentage; energy intake; and EE[i]	12 weeks-24 months	50% (6/12) of weight loss trials reported significantly greater weight loss among individuals randomized to technology interventions compared with controls; insufficient evidence to determine the effectiveness of interventions for weight maintenance.	Critically low
Dutton et al (2014) [55]	18 (22)	Narrative synthesis of all study types	Various: mobile, internet, and podcasts	Weight	3 weeks-24 months	67% (12/18) of trials found significant differences in weight loss at one or more assessments.	Critically low
Maxwell (2015) [64]	Not reported	Narrative synthesis of all study types	Technology interventions, including web-based and mobile	Healthy eating and active living	Not reported	Men participate in technology-based healthy lifestyle interventions less than women; maintenance of behavior is challenging.	Critically low
Podina and Fodor (2018) [66]	43 (47)	Meta-analysis of RCTs	Various: mobile messaging, mobile app, and website	Weight, BMI, waist circumference, and percentage of body fat	3-24 months	Standard active treatment was more effective than eHealth interventions with regard to weight (g=−0.31; 95% CI − 0.43 to −0.20). There was a statistically significant, albeit small effect size favoring eHealth interventions relative to passive control groups for weight (g=0.34; 95% CI 0.24 to 0.44) and behavioral outcomes (g=0.17; 95% CI 0.07 to 0.27).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Ryan et al (2019) [67]	6 (6)	Narrative synthesis of randomized trials	Various: mobile or internet	Weight loss	5 weeks-24 months	Tailored interventions were found to be more effective in supporting weight loss than generic or waitlist controls in 66% (4/6) of articles. Effect sizes were very small to moderate, with evidence of fluctuations in effect sizes and differences of effect between tailored and non-tailored interventions, and between tailoring types, over time.	Critically low
Nonactive controls							
An et al (2017) [49]	21 (22)	Meta-analysis of RCTs, pre-post studies, and cohort studies	Social media	Physical activity; sedentary behavior; diet; BMI; hip-waist ratio; body fat; and waist circumference	2-102 weeks	Social media–based interventions were found to reduce body weight by 1.01 kg (95% CI 0.45 to 1.57), BMI by 0.92 kg/m^2 (95% CI 0.29 to 1.54), and waist circumference by 2.65 cm (95% CI 0.86 to 4.43).	Critically low
Tang et al (2016) [72]	18 (27)	Meta-analysis of RCTs	Various	Body weight; BMI; and waist circumference	1-24 months	Participants receiving internet-based, self-directed interventions lost significantly more weight than those receiving minimal intervention or no treatment (MD −1.72 kg; 95% CI −2.60 to −0.84; significant mean difference −0.45; 95% CI −0.67 to −0.23; I^2=80%; *P*<.001) and a significantly greater reduction in BMI levels than those receiving no treatment or minimal intervention (MD −0.47 kg/m^2; 95% CI −0.81 to −0.14; significant mean difference −0.32; 95% CI −0.61 to −0.03; I^2=90%; *P*=.03; 13 evaluations). There was a greater reduction in BMI (MD 0.54 kg/m^2) and waist circumference (2.81 cm) at 0-4 months follow-up than at later times (k=27).	Critically low
Active controls							
Beleigoli et al (2019) [52]	11 (11)	Meta-analysis of RCTs	Internet	Weight and BMI	3-12 months	Compared with offline interventions, digital interventions led to a greater short-term (<6 months follow-up) weight loss (MD −2.13 kg; 95% CI −2.71 to −1.55; 393 participants; high-certainty evidence) but not in the long-term (MD −0.17 kg; 95% CI −2.10 to 1.76; 1104 participants; moderate-certainty evidence).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Kodama et al (2012) [58]	23 (23)	Meta-analysis of RCTs	Internet	Weight loss	3-30 months	Using the internet had a modest but significant additional weight loss effect compared with non-web user control groups (−0.68 kg; *P*=.03). Internet-based interventions were effective for weight loss (−1.00 kg; *P*<.001) but not a substitute for face-to-face support (+1.27 kg; *P*=.01). An additional effect on weight control was observed when the aim of using the internet was initial weight loss (−1.01 kg; *P*=.03) but was not observed when the aim was weight maintenance (+0.68 kg; *P*=.26); furthermore, it was effective to use the internet as an adjunct to face-to-face care (−1.00 kg; *P*<.001) but adverse effects on weight loss were found when it was used as a substitute (+1.27 kg; *P*=.01). The weight loss effect was insignificant (−0.20 kg; *P*=.75) in studies with educational periods ≥12 months and was significant in studies with an educational period <6 months (−1.55 kg; *P*=.001).	Critically Low

[a]RCT: randomized controlled trial.

[b]HbA_{1c}: glycated hemoglobin.

[c]LDL: low-density lipoprotein.

[d]MD: mean difference.

[e]SMD: standardized mean difference.

[f]PDA: personal digital assistant.

[g]BP: blood pressure.

[h]RHR: resting heart rate.

[i]EE: energy expenditure.

Internet interventions were found to be somewhat effective. A total of 5 narrative syntheses found them to be effective in approximately half of the studies [47,62,99,100,108]. In total, 3 meta-analyses quantified weight loss and other health effects. A total of 3 reviews found a significant reduction in weight: mean difference –1.34 kg [102], SMD 2.13 kg [70], and a 0.68-kg additional weight reduction over a period of 3 to 30 months [99]. A fourth review found that internet interventions significantly reduced waist circumference (mean change −2.99 cm) [69]. However, stratified analysis suggested that internet interventions were effective when used in combination with in-person counseling (−1.93 kg), rather than as a substitute [98]. This finding was supported by a narrative synthesis, which reported that internet interventions were more effective when they were enhanced to offer more than just educational resources (several studies found medium effect sizes) [62].

There were mixed results regarding the effectiveness of social media interventions. A total of 2 reviews were favorable [49,56], including a meta-analysis that found that social media–based interventions reduced body weight by 1.01 kg, BMI by 0.92 kg/m^2, and waist circumference by 2.65 cm but did not find significant changes in body fat or body fat percentage [49]. A total of 2 narrative syntheses found effects that were small and not meaningful [61,76]. Two meta-analyses found no significant effects on the diverse primary outcomes of the studies or on body weight [65,75].

Most reviews of mobile phone interventions found that they were effective in achieving weight loss via diet and physical activity. One meta-analysis found a statistically significant medium effect size (Cohen *d*=0.430) in favor of mobile phone interventions [60]. Two others quantified the change in terms of body weight: mobile phone interventions were associated with significant changes in body weight of −1.44 kg and in BMI of $-0.24\ kg/m^2$ compared with controls [59]; the weighted mean body weight change in intervention participants was −2.56 kg compared with −0.37 kg in controls [71]. These results were supported by 3 narrative syntheses, which found evidence that

XSL•FO
RenderX

mobile interventions led to weight loss [50,68,99]. However, one review of apps to promote healthy lifestyles found that most trials did not show significant differences [104], and one of the reviews that looked at the use of apps for diet, physical activity, and sedentary behavior, which concluded that apps led to weight loss, also concluded that the apps were more successful when used alongside other intervention components than when used alone [68].

In total, 4 of 5 narrative syntheses that ranged over a variety of digital interventions found that around half of the interventions were effective compared with controls [48,51,53,54], with only one finding that a large majority (81%) reported a significant reduction in adiposity [99]. One of these studies found that most trials reported within-group weight loss, even when there was no difference between digital interventions and controls [54]. Another study compared the success of different modes of delivery, finding the highest proportion of successful trials involving SMS text messaging or email (67%), followed by online chat rooms (50%), web-based (48%), and self-monitoring with technology (43%) [48].

Effectiveness of Digital Interventions on Diet and Physical Activity (Weight Loss) Compared With Nonactive Intervention Controls

A total of 6 reviews that reported results for digital interventions for diet and physical activity to no intervention or minimal intervention controls all agreed that digital interventions were more effective [66,67,69,70,72,74]. Table 5 provides a summary of the results.

Internet interventions were more effective than nonactive controls according to 3 meta-analyses and 1 narrative synthesis [67,69,70,72]. Internet interventions led to a greater reduction in waist circumference (mean change –2.99 cm, 95% CI −3.68 to −2.30, I2=93.3% vs 2.38 cm, 95% CI 1.61 to 3.25, I2=97.2%) [69]; internet-delivered personal feedback led to greater weight loss (mean difference 2.14 kg) [70]; and self-directed internet interventions led to significantly more weight loss (mean difference −1.56 kg) and showed a significantly greater reduction in BMI (mean difference −0.41 kg/m^2) [72].

Similar effects were observed for computer-based interventions, which led to greater weight loss at 6 months (mean difference −1.5 kg) and were superior to limiting weight regain (mean difference −0.7 kg) [74].

A narrative review of tailored internet interventions found that effect sizes ranged from very small to moderate [67]. This was supported by a meta-analysis that covered various digital interventions, which found small effect sizes favoring digital interventions for weight (g=0.34) and behavioral outcomes (g=0.17) [66].

Effectiveness of Digital Interventions on Diet and Physical Activity (Weight Loss) Compared With Active Controls

The effects of digital interventions on diet and physical activity (with regard to weight loss) compared with active controls were mixed. A total of 2 meta-analyses found no differences between web-based interventions and active offline interventions for weight loss outcomes [52,69]. One meta-analysis of a range of interventions found that standard active treatment was more effective for weight loss (g=0.31) [66]. However, one meta-analysis found that using the internet had a modest but significant additional weight loss effect compared with offline control groups (−0.68 kg; P=.03), with a subgroup analysis showing that the internet was effective compared with controls for achieving weight loss (weight change=−1.01 kg) but not weight maintenance [58]. The same meta-analysis also found that it was effective to use the internet as an adjunct to face-to-face care (−1 kg; P<.001) but that adverse effects on weight loss were found when it was used as a substitute (+1.27 kg; P=.01) [58].

Computer-based interventions led to less weight loss than in-person treatment (mean difference 2.1 kg) and were not superior to infrequent in-person treatment in limiting weight regain at 6 months [74].

Sustainability of Effects on Diet and Physical Activity (Weight Loss) at Follow-Up

Although technology-related health interventions may be effective, the maintenance of behavior is challenging [64], and there is insufficient evidence to determine the effectiveness of digital interventions on weight maintenance [54]. One review of internet interventions reported that, in studies finding significant weight reduction, the magnitude of weight change ranged from 1 to 6 kg after 6 months of follow-up [99].

Many studies included in the reviews had short follow-ups. Where longer follow-ups were reported, effectiveness typically diminished over time. Examples of the diminishing effects are clear in the 2 reviews. One meta-analysis of internet-delivered personal feedback found a greater reduction in BMI (mean difference 0.54 kg/m^2) and waist circumference (2.81 cm) at 0 to 4 months follow-up than at later times compared with undefined control groups [72]. A narrative synthesis supports these findings, concluding from multiple high-quality randomized controlled trials (RCTs) that weight loss occurs for a short term through mobile interventions, with moderate evidence for the medium term [50]. Two other meta-analyses reported significantly greater weight loss in favor of digital interventions in the medium term only (<6 months): 2.13 kg [52] and 1.55 kg [58]. One meta-analysis assessed differences between trials of different lengths, reporting no differences between shorter and longer trials (<6 or ≥6 months) [59].

Smoking

Review Characteristics

There were 28 reviews on smoking (see Multimedia Appendix 6 for summary characteristics).

Effectiveness of Digital Interventions on Smoking Compared With Mixed (Active and Nonactive) Controls

Digital interventions were generally effective across different modes of delivery (Table 6).

XSL•FO
RenderX

Table 6. Results of reviews on smoking, ordered by type of control and further ordered by mode of delivery of intervention.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed (active and nonactive) controls							
Afshin et al (2016) [99]	22 (224)	Narrative synthesis of RCTs[a] and quasiexperimental studies	Internet and mobile interventions	Abstinence	1 week-2 years	77% (17/22) of studies reported a significant increase in abstinence. In studies reporting benefits, the OR[b] for 7-day abstinence at 6 months ranged from 1.6 (95% CI 1.1 to 2.4) to 2.7 (95% CI 1.8 to 4.0).	Critically low
Boland et al (2016) [83] (Smoking only)	13 (13)	Meta-analysis of RCTs	Various: website or computer program and SMS text messaging	Abstinence	1 week-18 months	Interventions increased the odds of smoking cessation for disadvantaged groups at 1 month (OR 1.70; 95% CI 1.10 to 2.63), 3 months (OR 1.30; 95% CI 1.07 to 1.59), 6 months (OR 1.29, 95% CI 1.03 to 1.62) and 18 months postintervention (OR 1.83, 95% CI 1.11 to 3.01).	Low
Aneni et al (2014) [100]	3 (29)	Narrative synthesis of RCTs	Internet	Cessation	12 months	3 follow-up studies that measured smoking cessation showed significant intervention effects, although they were assessed to be of low quality.	Critically low
Chebli et al (2016) [103]	9 (16)	Narrative synthesis of RCTs	Internet	Cessation and reduction	1-12 months	Internet-based interventions may have a positive effect on smoking cessation. Several studies found that web-based use and number of log-ins was positively associated with quit outcomes.	Critically low
Cheung et al (2017) [85] (Smoking only)	6 (45)	Narrative synthesis and meta-analysis of RCTs and quasi-RCTs	Internet	Cessation	>4 weeks	Only 13% (6/45) of studies provided data on effectiveness, with 66% (4/6) of studies demonstrating effectiveness. Smokers using a web-based cessation intervention were 1.15 to 2.84 times more likely to become a former smoker compared with the control condition (with a pooled RR[c] 1.39; 95% CI 1.18 to 1.65).	Critically low
Gainsbury and Blaszczynski (2010) [87] (Smoking only)	7 (9)	Narrative synthesis of RCTs and pre-experimental studies	Internet	Abstinence, tobacco use, smoking status, compliance, nicotine dependence, carbon monoxide markers, and toxicity	3-12 months	86% (6/7) of studies reported significantly greater self-reported smoking quit rates or abstinence at the end of the treatment trial for participants in the internet intervention compared with controls. Several trials found improvements at 3, 6, and 12 months.	Critically low
Graham et al (2016) [88] (Smoking only)	40 (40)	Meta-analysis of RCTs	Internet	Abstinence	7 days-3 months	Pooled results from 15 trials (24 comparisons) found a significant effect in favor of experimental internet interventions (RR 1.16; 95% CI 1.03 to 1.31; I^2=76.7%).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Hutton et al (2011) [90] (Smoking only)	15 (21)	Narrative synthesis of RCTs	Internet	Cessation	>1 month	Two RCTs found that a multicomponent intervention with web and nonweb-based elements was more efficacious than a self-help manual, and one of the 2 RCTs found that web-based interventions may be more effective than no treatment. Three trials provided insufficient evidence to demonstrate whether web-based interventions were more efficacious than counselling. Tailored websites in 2 RCTs and greater website exposure in 86% (6/7) of RCTs were associated with higher rates of abstinence.	Critically low
Lustria et al (2013) [110]	8 (40)	Meta-analysis of experimental and quasiexperimental studies	Internet	Abstinence	30 days-6 months	Web ‐ based, tailored interventions had significantly greater improvement in smoking outcomes compared with control conditions, with small effects, Cohen *d*=0.151 (k=8; 95% CI 0.11 to 0.19; *P*<.001).	Critically low
McCrabb et al (2019) [91] (Smoking only)	45 (45)	Meta-analysis of RCTs	Internet	Abstinence	1-18 months	Interventions were effective in the short term (OR 1.29, 95% CI 1.12 to 1.50; *P*=.001) and long term (OR 1.1.9, 95% CI 1.06 to 1.35; *P*=.004).	Low
Shahab and McEwen (2009) [95] (Smoking only)	10 (11)	Meta-analysis of RCTs	Internet	Cessation	>1 month	Interactive interventions were effective compared with untailored booklets or emails (RR 1.8; 95% CI 1.4 to 2.3) increasing 6-month abstinence by 17% (95% CI 12 to 21%); no evidence was found of a difference between interactive and static interventions.	Critically low
Taylor et al (2017) [97] (Smoking only)	61 (67)	Meta-analysis of RCTs and quasi-RCTs	Internet	Cessation	6-12 months	Interactive and tailored internet-based interventions with or without additional behavioral support are moderately more effective than nonactive controls at 6 months or longer, but there was no evidence that these interventions were better than other active smoking treatments.	Moderate
Webb et al (2010) [113]	12 (85)	Meta-analysis of RCTs	Internet	Smoking abstinence	12 months	Interventions that targeted smoking abstinence tended to have small effects on behavior that did not reach statistical significance (Cohen *d*+=0.07; k=12; 95% CI −0.04 to 0.18).	Critically low
Head et al (2013) [107]	5 (19)	Meta-analysis of RCTs	SMS text messaging	Smoking cessation	Mean 81.26 days	The weighted mean effect size for smoking cessation, Cohen *d*=0.447 (95% CI .367 to .526; *P*=.001; k=5).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Scott-Sheldon et al (2016) [94] (smoking only)	18 (20)	Meta-analysis of RCTs	SMS text messaging	Abstinence, cigarette use, quit attempts, and nicotine dependence	Not reported	SMS text messaging was associated with significantly greater odds of abstinence compared with controls: 7-day point prevalence (OR 1.38, 95% CI 1.22 to 1.55; k=16) and continuous abstinence (OR 1.63, 95% CI 1.19 to 2.24; k=7); interventions were also more successful in reducing cigarette consumption (Cohen d_+=0.14; 95% CI 0.05 to 0.23; k=9).	Critically low
Spohr et al (2015) [96] (smoking only)	10 (13)	Meta-analysis of RCTs	SMS text messaging	Cessation	3 and 6 months	Interventions generally increased quit rates compared with controls (OR 1.36, 95% CI 1.23 to 1.51). Intervention efficacy was higher in studies with a 3-month follow-up compared with a 6-month follow-up.	Critically low
Palmer et al (2018) [111]	18 (71)	Meta-analysis of RCTs	Mobile	Abstinence and cessation (verified biochemically)	24 hours-6 months	The effect of SMS text messaging–based smoking cessation support on biochemically verified continuous abstinence was pooled relative risk, RR 2.19 (95% CI 1.80 to 2.68; I^2=0%) and on verified 7-day point prevalence of smoking cessation was pooled RR 1.51 (95% CI 1.06 to 2.15; I^2=0%).	Moderate
Whittaker et al (2016) [98] (smoking only)	12 (12)	Meta-analysis of RCTs and quasi-RCTs	Mobile	Cessation	6 months	Smokers who received support programs were 1.7 times more likely to stay quit than smokers who did not receive the programs (9.3% quit with programs compared with 5.6% who quit with no programs). Most of the studies were of programs relying mainly on text messages.	Moderate
Danielsson et al (2014) [86] (smoking only)	21 (74)	Narrative synthesis of RCTs	Various	Abstinence	>3 months	The studies showed mixed results regarding internet interventions and smoking, with some positive effects for the smoking cessation program that combined the use of both the internet, mobile phones (SMS text messaging), and email.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
HIQA (2017) [89] (smoking only)	12 (143)[d]	Network meta-analyses of RCTs	Various: internet and mobile	Cessation	6-12 months	Internet-based interventions are superior to control (brief advice or written materials; RR 1.43, 95% CI 1.02 to 2.00; *P*=.04; k=5); Internet-based interventions are superior to doing nothing (RR 1.46, 95% CI 1.18 to 1.81; *P*<.001; k=3); mobile phone–based interventions appear to have similar effectiveness to control (RR 1.18, 95% CI 0.88 to 1.60; *P*=.27; k=3); no evidence of difference between mobile phone–based interventions and internet (RR 1.43, 95% CI 0.88 to 2.31; *P*=.15; k=1).	Low
Hou et al (2013) [108]	5 (38)	Narrative synthesis of studies with comparison or control groups	Web-based computer programs	Cessation	Not reported	2 studies found higher cessation rates in intervention groups than control. 3 studies found no significant differences in quit rates at the end of the intervention or at follow-ups.	Critically low
Myung et al (2009) [92] (smoking only)	22 (22)	Meta-analysis of RCTs	Various: internet or computer based	Abstinence and biochemical markers	>3 months	Intervention groups had a significant effect on smoking cessation (RR 1.44; 95% CI 1.27 to 1.64). Similar findings were observed in web-based interventions (RR 1.40; 95% CI 1.13 to 1.72) and in computer-based interventions (RR 1.48; 95% CI 1.25 to 1.76).	Critically low
Naslund et al (2017) [93] (smoking only)	7 (7)	Narrative synthesis of all study types	Social media	Cessation	30-365 days	71% (5/7) of studies reported significant effects on smoking-related outcomes such as greater abstinence, reduction in relapse, and an increase in quit attempts.	Critically low
Elaheebocus et al (2018) [56]	7 (134)	Narrative synthesis of RCTs	Social media	Cessation	6-48 months	100% (7/7) of studies on smoking cessation using web-based social networks had positive results.	Critically low
Rooke et al (2010) [112]	13 (34)	Meta-analysis of RCTs	Computer based	Abstinence and reduction	1-156 weeks	The weighted average effect size (Cohen *d*) was 0.14; *P*<.001 for studies addressing tobacco use.	Critically low
Nonactive controls							
Chen et al (2012) [84] (smoking only)	60 (60)	Meta-analysis of RCTs and quasi-RCTs	Various: computer and other electronic aids	Cessation	2 days-30 months	Computer and other electronic aids increase the likelihood of cessation compared with no intervention or generic self-help materials but the effect is small (prolonged abstinence: RR 1.32; 95% CI 1.21 to 1.45).	Low
Krebs et al (2010) [109]	32 (88)	Meta-analysis of RCTs	Computer tailored	Abstinence	24 hours-9 months	Mean effect for the 32 studies reporting point prevalence outcome was *g*=0.16 (95% CI 0.12 to 0.19); mean effect for the 16 studies reporting prolonged abstinence measures was *g*=0.24 (95% CI 0.20 to 0.31).	Low
Active controls							

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Covolo et al (2017) [104]	2 (40)	Narrative synthesis of RCTs	Mobile apps	30-day point prevalence cessation	30 days	1 trial compared 2 apps and found no evidence of any difference; the other found text messaging produced more abstinence than an app (P<.05).	Critically low

[a]RCT: randomized controlled trial.

[b]OR: odds ratio.

[c]RR: risk ratio.

[d]In total, 12 relevant studies were included in the meta-analysis.

Internet interventions were generally found to be effective (Table 6). A total of 3 relevant meta-analyses found small or small-to-medium effects [85,89,91]: pooled RR 1.39 [85]; RR 1.40 [92], and RR 1.43 [88] and a short-term (≤6 months) OR 1.29 (95% CI 1.12 to 1.50; P=.001) [91]; the latter showed that internet interventions were also successful for the individual outcomes of *prolonged abstinence*, that is, not smoking since a quit date (OR 1.43) and *30-day point prevalence abstinence*, that is, not smoking one or more days before the follow-up (OR 1.75) [91]. However, one narrative synthesis noted that although the 3 studies that measured smoking cessation all showed significant intervention effects, they were all assessed to be of low quality [100].

The 3 reviews of mobile interventions that focused on SMS text messaging found it to be effective, again with small effects: Cohen d=0.14 [94]; OR 1.36 [96]; and OR 1.36 [98]. A fourth meta-analysis differentiated between specific outcomes and found a medium effect (RR 2.19) on biochemically verified continuous abstinence, and a small-to-medium effect (RR 1.51) on verified 7-day point prevalence of smoking cessation was pooled [111]. However, 2 reviews (including 1 meta-analysis) with information on mobile interventions in general did not find that they were effective [89,99].

A meta-analysis of computer-delivered interventions also found a small but significant effect size associated with studies addressing tobacco use (Cohen d=0.14) [112].

There is mixed evidence from reviews that cover a variety of digital technologies. One narrative synthesis [99] and one meta-analysis found that they were effective, with interventions increasing the odds of smoking cessation at 1 month (OR 1.70) [83]. However, another narrative synthesis that included both phone and internet interventions found that there were mixed results [86].

A total of 2 reviews of social media interventions with narrative syntheses both found favorable effects of the interventions on outcomes related to smoking cessation [56,93].

As seen above, separate reviews of internet, mobile, and computer-delivered interventions found small effects for each. One meta-analysis looked for differences in effect sizes between different modes of intervention and found no statistically significant evidence of any differences in effect sizes of internet interventions, intensive advice, telephone support, individual counseling, or group behavior therapy [89].

Effectiveness of Digital Interventions on Smoking Compared With Nonactive Controls

There were small effects of digital interventions compared with nonactive controls for all modes of delivery (Table 6). Meta-analyses found small effects for pooled modes of digital interventions compared with controls: 1 month OR 1.70 (k=13) [83]; pooled estimate for prolonged abstinence, RR 1.32 (k=60) [84], and pooled estimate for point prevalence abstinence, RR 1.14 [84]. Internet interventions were more effective than nonactive controls [88-90,92,97], with 3 meta-analyses finding small effects compared with nonactive controls: RR 1.15 [97], RR 1.46 [89], and RR 1.60 [88]. Computer-tailored interventions had effects sizes of g=0.16 for point prevalence outcomes and g=0.24 for prolonged abstinence [109]. Mobile phone–based interventions appeared to have similar effectiveness to minimal controls (RR 1.18) and were more effective than no intervention (typically either waiting list control or no further contact until follow-up) [89].

Effectiveness of Digital Interventions on Smoking Compared With Active Controls

Most reviews concluded that internet interventions were not more effective than active controls, with 3 reviews (including 2 meta-analyses) not finding differences [88,90,97]. This included a meta-analysis that found no significant effects of internet interventions compared with face-to-face counseling or telephone counseling [88]. However, one narrative synthesis was more positive about the greater effect of internet interventions compared with conventional ones [99].

There was no information on mobile interventions compared with active controls, but one meta-analysis with mixed controls noted that the summary effect sizes favored the treatment groups even when 18 of the 20 (90%) controlled trials used an active control and 12 (66%) of these active controls included some smoking-related content, including smoke-free websites, self-help guidebooks, and smartphone apps [94].

Sustainability of Effects on Smoking at Follow-Up

There was agreement that the effects of internet interventions were sustained for up to 6 months [83,87,91,95]. This was quantified as interventions increasing the 6-month abstinence by 17% [95], increasing the odds of cessation at the 6-month

XSL•FO
RenderX

follow-up (OR 1.29) [83], and increasing the likelihood of *prolonged abstinence* (ie, not smoking since a quit date; OR 1.19) [91]. However, there was disagreement about whether there were significant effects at 12-month follow-ups, with 2 reviews in favor [87,89], one quantifying the quit rate at 12 months as 8% [89] but another reporting that the positive results of internet interventions were not generally maintained at 1-year follow-up assessments [103].

There is mixed evidence on the sustainability of the effects of mobile interventions. A total of 2 meta-analyses found that mobile interventions were effective in the medium term, with a quit rate of 13% [89] and effects on biochemical measures of quitting at 6 months (RR 1.83) [98]. However, a study of SMS text messages suggested that any effect was not sustained [96].

Regarding computer-delivered interventions, one meta-analysis found that computer programs increased the odds of cessation at the 3-month (OR 2.04), 12-month (OR 1.68), and 18-month follow-up (OR 1.83) [83]. A meta-analysis of computer-delivered interventions did not find evidence of differential effects depending on the length of follow-up or number of sessions [112].

A total of 2 reviews that considered a range of different technologies found evidence of sustained effects. A narrative synthesis of internet and mobile interventions found that the OR for 7-day abstinence at 6 months ranged from 1.6 to 2.7 [99]. A meta-analysis found that digital interventions increased the odds of smoking cessation at 3 months (OR 1.30), 6 months (OR 1.29), and 18 months (OR 1.83) postintervention [83].

In a pooled analysis of web-based and computer-based interventions, the smoking cessation rate was 14.8% in the intervention group and 14.3% in the control group at the short-term 3-month follow-up (P=.42), 11.7% and 7.0% at the midterm 6- to 10-month follow-up (P<.001), and 9.9% and 5.7% at the long-term 12-month follow-up (P<.001), respectively [92].

One meta-analysis that covered multiple intervention types found that SMS text messaging had the highest OR (2.81) at the 1-month follow-up, followed by the use of computer programs at the 3-month follow-up (OR 2.04), 12-month follow-up (OR 1.68), and 18-month follow-up (OR 1.83) and websites at 6 months (OR 1.37), while a DVD intervention and integrated videotelephony did not increase the odds of cessation compared with no intervention [83].

Alcohol

A total of 13 papers covered alcohol [73,77-82,99,103,104,111-113]. Multimedia Appendix 5 presents a breakdown of the study characteristics.

Effectiveness of Digital Interventions on Alcohol Consumption Compared With Mixed (Active and Nonactive) Controls

In total, 7 of 8 (88%) reviews with mixed controls reported favorable evidence that digital interventions were effective in decreasing alcohol consumption (Table 7), although one review noted that the controls were just as effective [103]. A total of 2 meta-analyses reported small effect sizes (Cohen d=0.26 [112] and Cohen d=0.14 [113]). A narrative synthesis of interventions using novel technologies found that, in studies finding benefits and reporting compliance with drinking recommendation as an outcome, the OR for drinking within the recommended limit ranged from 1.7 to 3.7 (small/medium to medium/large effects) [99]. In contrast, the sixth review, which surveyed mobile phone interventions, reported that the results were inconclusive for alcohol reduction, and the authors declined to perform a meta-analysis because the results were self-reported and therefore at risk of overstating the benefits [111].

XSL•FO
RenderX

Table 7. Effectiveness of digital interventions on alcohol consumption, sorted by controls and further ordered by intervention type.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed (active and nonactive) controls							
Afshin et al (2016) [99]	47 (224)	Narrative synthesis of RCTs[a] and quasiexperimental studies	Various: internet, mobile, computer software, and sensors	Alcohol frequency and quantity, binge drinking, estimated blood alcohol concentration, alcohol dependency, and Alcohol Use Disorder Identification Test scores	1 week-2 years	34% (39/47) of studies, 41 RCTs and 6 quasiexperimental, reported a significant decrease in alcohol use; 83% (33/41) of RCTs reported statistically significant benefits.	Critically low
Chebli et al (2016) [103]	2 (16)	Narrative synthesis of RCTs	Internet	Cessation and reduction of alcohol	1-12 months	Both studies demonstrated positive treatment outcomes in both arms, but there were no differences between internet intervention and control.	Critically low
Kaner et al (2017) [77] (Alcohol only)	57 (57)	Narrative synthesis and meta-analysis of RCTs	Computer and mobile	Alcohol consumption and frequency	1-12 months	Alcohol consumption reduced by approximately 23 g per week (95% CI 15 to 30) at follow-up (1-12 months; based on 41 studies). Frequency of consumption reduced (based on 15 studies): participants who engaged with digital interventions had less than one drinking day per month fewer than no intervention controls (moderate - quality evidence); had about one binge drinking session less per month in the intervention group (moderate - quality evidence); and drank one unit per occasion less than no intervention control participants (moderate - quality evidence). Compared with face-to-face interventions, there was no difference in alcohol consumption at the end of follow-up (mean difference 0.52 g/week; 95% CI −24.59 to 25.63; low - quality evidence).	Low
Kolar et al (2015) [78] (Alcohol only)	2 (18)	Narrative synthesis of all studies	Internet	Alcohol quantity and frequency	1 month	100% (2/2) of studies found reduced alcohol consumption in both arms but no significant differences between arms.	Critically low
Palmer et al (2018) [111]	8 (71)	Narrative synthesis of RCTs	Mobile	Self-report alcohol consumption	Not reported	The effects of alcohol reduction interventions were inconclusive.	Moderate
Rooke et al (2010) [112]	9 (34)	Meta-analysis of RCTs	Computer-delivered	Abstinence and reduction of alcohol	1-156 weeks	The weighted average effect size (Cohen *d*) was 0.20 (*P*<.001).	Critically low
Vernon et al (2010) [81] (Alcohol only)	15 (15)	Narrative synthesis of all studies	Computer-delivered	Alcohol consumption	30 days-12 months	All but one intervention showed significant improvement in at least one drinking-related outcome. However, interventions were heterogenous and preintervention alcohol consumption was not standardized.	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Webb et al (2010) [113]	9 (85)	Meta-analysis of RCTs	Internet	Alcohol consumption	Not reported	Small effects were observed for alcohol consumption (Cohen d_+=0.14; k=9; 95% CI 0.00 to 0.27).	Critically low
Nonactive controls							
Black et al (2016) [82] (Alcohol only)	93 (93)	Meta-analysis of RCTs	Computer delivered	Alcohol consumption: total consumption over a period of time; average alcohol consumption per drinking occasion or drinking day; peak consumption—max consumed on one occasion. Frequency of heavy episodic drinking and of any alcohol consumption	Up to 2 years	Small effects averaging across timepoints, Cohen *d*=0.007 (heavy episodic drinking frequency) to Cohen *d*=0.15 (total consumption); in the short term, there were small-to-medium effects (Cohen *d*+=0.16 to 0.31) and significant effects on all outcomes except drinking frequency; in the medium-to-long term, they produced small (Cohen *d*+=0.07 to 0.12), significant effects on all outcomes.	Critically low
Covolo et al (2017) [104]	1 (40)	Narrative synthesis of RCTs	Mobile	Alcohol frequency	0-2 years	Contrary to expectation, it was found that the mobile app significantly increased the frequency of drinking occasions compared with the control group (*P*=.001).	Critically low
Riper et al (2011) [79] (Alcohol only)	9 (9)	Meta-analysis of RCTs	Internet	Alcohol consumption	Up to 12 months	An overall medium effect size (*g*=0.44; 95% CI 0.17 to 0.71; random effect model) in favor of the intervention groups was found.	Critically low
Riper et al (2014) [80] (Alcohol only)	16 (16)	Meta-analysis of RCTs	Internet	Alcohol consumption	1 to 12 months	A small but significant overall effect size in favor of internet interventions (*g*=0.20; 95% CI 0.13 to 0.27; *P*=.001) was found. Participants in internet-based interventions consumed approximately 22 g of ethanol less and were more likely to adhere to low-risk drinking guidelines (risk difference 0.13; 95% CI 0.09 to 0.17; *P*=.001).	Critically low
Tsoli et al (2018) [73]	4 (15)	Meta-analysis of RCTs	Interactive voice responses	Alcohol consumption	6 weeks-12 months	The meta-analysis of included studies demonstrated that interactive voice response–based interventions had no statistically significant effect on alcohol consumption (*g*=−0.077; 95% CI −0.162 to 0.007; k=4; *P*=.07).	Critically low

[a]RCT: randomized controlled trial.

Effectiveness of Digital Interventions on Alcohol Consumption Compared With Nonactive Controls

In total, 4 of 6 (67%) reviews with nonactive controls found an effect of the intervention (Table 7). Of the 4 meta-analyses, 2 found a medium effect on alcohol consumption (*g*=0.44 [80] and *g*=0.44 [79]). A third found a small-to-medium effect on total alcohol consumption (Cohen *d*+=0.31), small effects on 2 other consumption measures, and a measure of frequency of heavy drinking episodes (Cohen *d*+=0.16-0.19) and no effect on drinking frequency [82]. The fourth review found no effect [73].

Two studies quantified the effect on alcohol consumption, finding a reduction in weekly consumption of 22 g of alcohol [80] and 23 g of alcohol [77], approximately 3 UK units. However, when studies with a high risk of bias were excluded, this number decreased to 11 g of alcohol (or 1.5 UK units) per week [77]. With regard to frequency of consumption, participants who were given digital interventions had less than one drinking day per month, which is fewer compared with

XSL•FO
RenderX

no-intervention controls (moderate-quality evidence); had approximately 1 binge-drinking session less per month in the intervention group compared with no-intervention controls (moderate-quality evidence); and drank 1 unit per occasion, which is less than the no-intervention control participants (moderate-quality evidence) [77]. Participants in internet interventions were significantly more likely to adhere to low-risk drinking guidelines at the immediate posttreatment follow-up, compared with the no-intervention controls [80].

Only one review was negative. A narrative synthesis of mobile apps reported that there was a single RCT on alcohol reduction where, contrary to expectation, the frequency of drinking occasions was higher in the intervention group [104].

Effectiveness of Digital Interventions on Alcohol Consumption Compared With Active Controls

In total, 2 reviews that separated active controls found that there were no significant differences between the intervention and control groups [77,99], one of which was specifically compared with face-to-face controls (Table 7) [77].

Sustainability of Effects on Alcohol Consumption at Follow-Up

The effectiveness of the interventions seemed to decrease over time (Table 7). One meta-analysis, which reported small-to-medium effects in the short term (Cohen *d*+=0.16-0.19), found that in the medium to long term, there were small (Cohen *d*+=0.07-0.12), significant effects on all outcomes [82]. A total of 2 reviews reported that, for internet-based intervention studies where there were 3-, 6-, or 12-month follow-up data, no significant differences in effect remained in later follow-up [80,103]. In contrast, one review of computer- or mobile-delivered interventions found that positive differences in measures of drinking were seen at 1, 6, and 12 months [77], and one review of internet interventions found a medium effect size (*g*=0.39), lasting up to 6 or 9 months posttreatment, as compared with no intervention; the effects of the interventions beyond 9 months could not be assessed, but 2 studies in the review suggested that they had faded out by 12 months [79].

Other Combinations

Review Characteristics

A total of 11 reviews covered a number of areas of behavior [101-107,109,110,112,113]. The breakdown of their characteristics is shown in Multimedia Appendix 5.

Effectiveness of Other Digital Combination Interventions Compared With Mixed (Active and Nonactive) Controls

All 5 reviews of internet interventions concluded that they were effective in changing behavior (Table 8). A total of 3 meta-analyses found small effects: Cohen *d*=0.19 [105], Cohen *d*=0.139 [110], and Cohen *d*=0.16 [113]. However, one study reported that the effect sizes were heterogeneous [113]. The fourth meta-analysis quantified the effects on health outcomes, finding statistically significant reductions in systolic blood pressure (mean difference –2.66 mm Hg), diastolic blood pressure (mean difference –1.26 mm Hg), glycated hemoglobin level (mean difference –0.13%), and low-density lipoprotein cholesterol level (mean difference –2.18 mg/dL) [102].

XSL•FO
RenderX

Table 8. Effectiveness of digital interventions on other combinations of outcomes, sorted by controls and further sorted by intervention type.

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Mixed (active and nonactive) controls							
Beishuizen et al (2016) [102]	15 (57)	Meta-analysis of RCTs[a]	Internet	Systolic blood pressure, diastolic blood pressure, HbA_{1c}[b] level, cholesterol level, weight, and level of physical activity	3-60 months	Intervention groups had a reduction in systolic blood pressure (mean difference –2.66 mm Hg; 95% CI –3.81 to –1.52), diastolic blood pressure (mean difference –1.26 mm Hg; 95% CI –1.92 to –0.60), HbA_{1c} level (mean difference –0.13%; 95% CI –0.22 to –0.05), and LDL[c] cholesterol level (mean difference –2.18 mg/dL; 95% CI –3.96 to –0.41). There were larger effects in internet interventions that combined the internet application with human support (blended care).	Low
Chebli et al (2016) [103]	11 (16)	Narrative synthesis of RCTs	Internet	Cessation and reduction (for both alcohol and smoking)	1-12 months	Internet-based interventions may have a positive effect on smoking cessation. Several studies found that web-based use and number of logins were positively associated with quit outcomes. Both studies on alcohol demonstrated positive treatment outcomes in both arms, but there were no differences between the internet-based intervention and control.	Critically low
Cugelman et al (2011) [105]	30 (30)	Meta-analysis of experimental, quasiexperimental, and correlational studies	Internet	Behavioral outcomes	1 day-7 months	Effect sizes were small but statistically significant (standardized mean difference effect size Cohen *d*=0.19; 95% CI 0.11 to 0.28; *P*<.001; number of interventions, k=30); however, there was a lot of heterogeneity, Cochran's Q test 64.125 (*P*<.001) and I^2=54.776; the largest effect size was observed when interventions were compared with waitlists and placebos (Cohen *d*=0.28; 95% CI 0.17 to 0.39; *P*<.001; k=18); there was no significant difference compared with sophisticated print interventions (Cohen *d*=–0.11; *P*>.05; k=29).	Critically low
Webb et al (2010) [113]	85 (85)	Meta-analysis of RCTs	Internet	Smoking abstinence, level of physical activity, alcohol consumption, and dietary behavior	3-12 months	Interventions had a statistically small but significant effect on health-related behavior (Cohen *d*+=0.16; 95% CI 0.09 to 0.23). The effect size of interventions targeting a single area of health behavior was not significantly different to the effect size of those targeting multiple areas of health behaviors, but the numerical difference was in favor of single-area studies (Cohen *d*+=0.17 versus Cohen *d*+=0.12).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Lustria et al (2013) [110]	40 (40)	Meta-analysis of experimental and quasiexperimental studies	Tailored internet-based interventions	Levels of physical activity, fruit and vegetable intake, saturated fat intake, and abstinence from smoking and alcohol consumption	1-24 months	Web ‐ based, tailored interventions effected significantly greater improvement in health outcomes as compared with control conditions immediately post intervention, Cohen *d*=0.139 (95% CI 0.111 to 0.166; *P*<.001; k=40). The effect remained at follow ‐ up, Cohen *d*=0.158 (95% CI 0.124 to 0.192; *P*<.001; k=21), and the correlation between follow ‐ up time point and effect size was r_{39}=.004 (*P*=.98, for posttest effects) and r_{20}=−.176 (*P*=.50, for follow ‐ up effects), which suggests that length of follow ‐ up did not significantly influence intervention outcomes. Interventions using tailored websites had a larger weighted mean effect size when compared with nontailored websites (Cohen *d*=0.188) than when they were compared with no ‐ treatment control conditions (Cohen *d*=0.07; *P*<.01). There was an extremely small effect of tailored websites frompared to no-treatment control conditions (Cohen *d*=0.08; k=4) and of tailored websites compared with sophisticated print interventions (Cohen *d*=–0.11; *P*>.05; k=29).	Critically low
Covolo et al (2017) [104]	39 (40)	Narrative synthesis of RCTs	Mobile apps	BMI and waist circumference; various physical activity levels; fruit and vegetable intake, high-sugar food intake, smoking cessation, and number of drinking days	6 months-2 years	Only 25% (10/40) of RCTs found statistical differences between intervention and control groups.	Critically low
Armanasco et al (2017) [101]	35 (51)	Meta-analysis of RCTs and quasiexperimental studies	Text messaging	Weight, level of physical activity, and smoking cessation	1 to 66 weeks	The overall pooled effect of interventions was Cohen *d*=0.24 (95% CI 0.16 to 0.32; *P*<.001; k=35) using outcome data collected most proximal to the intervention end. 7 studies collected data following a no-intervention maintenance period and showed a small but significant pooled maintenance effect (Cohen *d*=0.17; 95% CI 0.03 to 0.31; *P*=.017; k=7).	Critically low

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Head et al (2013) [107]	12 (19)	Meta-analysis of RCTs	Text messaging	Smoking cessation, weight loss, and level of physical activity	Mean 81.26 days	The overall weighted mean effect size representing the impact of these interventions on health outcomes was Cohen *d*=0.329 (95% CI 0.274 to 0.385; *P*<.001). Correlations between effect size and follow-up (r_{18}=−.12; *P*=.62) and effect size and retention (r_{18}=.14; *P*=.56) were not statistically significant. Effect sizes for interventions that employed no-treatment control groups (Cohen *d*=0.369), however, were significantly larger than those that employed alternative comparisons (Cohen *d*=0.226; Q_B=.5.16, df=1; *P*=.02).	Critically low
De Leon et al (2014) [106]	55 (55)	Narrative synthesis of RCTs	Text messages and emails (prompts)	Smoking cessation, diet intake, and physical activity level	3 weeks-9 months	76% (42/55) of articles found statistically significant positive behavioral outcomes of prompts, the mode by which the prompt was sent did not seem to impact its success.	Critically low
Rooke et al (2010) [112]	34 (34)	Meta-analysis of RCTs	Computer delivered	Abstinence and reduction for both smoking and alcohol	1-156 weeks	The weighted average effect size (Cohen *d*) was 0.20 (*P*<.001); however, lower effect sizes were associated with studies addressing tobacco use (Cohen *d*=0.14). There was significant heterogeneity between studies targeting tobacco versus alcohol use (*Q*=5.65; *P*=.02), with studies on alcohol consumption producing significantly higher effect sizes than tobacco studies (Cohen *d*=0.26 and 0.12, respectively). Effect sizes were higher for studies in which the comparison condition was an attention/placebo (Cohen *d*=0.22) relative to studies in which the comparison condition was an active comparison (Cohen *d*=0.10); studies employing active treatments as the comparison condition mainly produced effect sizes close to zero.	Critically low
Nonactive controls							

XSL•FO
RenderX

Review	Relevant studies, n (total studies)	Method of synthesis	Interventions	Outcomes	Follow-up	Summary of findings	AMSTAR-2 rating
Krebs et al (2010) [109]	76 (88)	Meta-analysis of RCTs	Computer-tailored interventions	Dietary intake of fat, vegetables, fruit; various levels of physical activity, and smoking abstinence over various timescales	1-24 months	The overall effect size was *g*=0.17 (95% CI 0.14 to 0.19) using a fixed effects model and *g*=0.17 using a random effects model (95% CI 0.14 to 0.20). Effects peak from 4 to 12 months postbaseline with a mean effect size of *g*=0.20, and while they decline after 12 months postbaseline, the mean effect size at long-term follow-up (*g*=0.12) remains to be statistically significant (95% CI 0.08 to 0.16). The meta-analysis found a trend for increasing effect sizes across studies that intervened on 1 (*g*=0.15), 2 (*g*=0.21), and 3 (*g*=0.24) areas of behavior, but this trend did not continue with the 1 study that intervened on 4 areas (*g*=0.12).	Low

[a]RCT: randomized controlled trial.

[b]HbA_{1c}: glycated hemoglobin.

[c]LDL: low-density lipoprotein.

Of the 3 reviews of mobile interventions, a review of SMS text messaging concluded that they were effective, but the narrative synthesis of apps was the only review in the other combinations category that did not find a preponderance of positive results [104]. The effect sizes for the SMS text messaging interventions were small (Cohen *d*=0.24 [101] and Cohen *d*=0.329 [107]), and the effect sizes in the meta-analysis were heterogeneous [107].

Other modes of intervention continued in the pattern of positive, albeit small, effects. Computer-delivered interventions had a small effect (Cohen *d*=0.20) [112]. Prompts delivered by SMS text messaging or email were effective in changing diet, physical activity, and smoking behaviors [106].

Relative Effect of Interventions by Number of Behaviors Targeted

A total of 3 meta-analyses that compared interventions targeting single versus multiple areas of behaviors found that both types of interventions were effective but they differed in whether they found interventions with single or multiple targets to be more effective. A total of 2 meta-analyses of internet-based interventions found that the effect size of interventions targeting a single area of health behavior was not significantly different to the effect size of those targeting multiple areas of health behaviors, but the numerical difference was in favor of single-area studies (Cohen *d*=0.146 vs Cohen *d*=0.121 for multiple-behavior studies [110]; Cohen *d*+=0.17 vs Cohen *d*+=0.12 [113]). However, the computer feedback–based meta-analysis found a trend for increased effect sizes across studies that intervened in 1 (*g*=0.15), 2 (*g*=0.21), and 3 (*g*=0.24) areas of behavior, but this trend did not continue with the one study that intervened in 4 areas (*g*=0.12) [109]. The review covered diet, physical activity, smoking cessation, and mammography.

Relative Effect of Intervention by Area of Behaviors Targeted

The review of SMS text messaging interventions found that interventions targeting smoking cessation and physical activity were more successful than interventions targeting other areas of health behavior, including alcohol and weight loss [107]. In contrast, a review of computer-delivered interventions found that interventions targeting smoking had statistically significantly lower effect sizes than those targeting alcohol (Cohen *d*=0.26 and 0.12, respectively) [112].

Effect of Other Combination Interventions Compared With Nonactive Controls

Digital interventions had their largest effect sizes when compared with nonactive controls; however, the effect sizes were generally small. Internet interventions had the largest effect size when compared with waitlists and placebos, but the effect sizes were small: Cohen *d*=0.28 [105] and Cohen *d*=0.22 [107]. Tailored websites had extremely small effect sizes when compared with no-treatment control conditions (Cohen *d*=0.07), and the effects were not statistically significant compared with nontailored print materials [110]. There were medium effect sizes for SMS text messaging interventions that employed no-treatment control groups (Cohen *d*=0.369) [107].

Effect of Other Combination Interventions Compared With Active Controls

The effects of digital interventions compared with active controls were very small or nonexistent.

A total of 2 meta-analyses found that there was an extremely small effect size (Cohen *d*=0.08) [110] or no significant difference [105] when internet-based interventions were compared with sophisticated print interventions. There were larger effects in internet-based interventions that combined

XSL•FO
RenderX

internet application with human support (blended care) [102]. Computer-delivered interventions had very small effect sizes when the comparison condition was an active comparison (Cohen *d*=0.10); studies employing active treatments as the comparison condition mainly produced effect sizes close to zero [112]. SMS text messaging interventions had small effect sizes compared with active comparisons (Cohen *d*=0.226), which was statistically significantly smaller than their effect size compared with the no-treatment control groups [107].

Although digital feedback was effective, it seems that it was no more effective than feedback delivered by nondigital means, and 2 reviews that examined feedback interventions found that the medium did not affect behavior change (SMS text messaging, print, email, telephone, newspaper articles [106] and print, computer, telephone, etc [109]).

Sustainability of Effects at Follow-Up

There was mixed evidence on the sustainability of interventions. A meta-analysis of internet-based interventions found the largest effect size at 1 month to 4 months [105], but a meta-analysis of computer-tailored interventions found that effects peaked from 4 months to 12 months postbaseline [109]. A total of 3 meta-analyses found that the correlation between effect size and follow-up was not significant for internet-based [110], SMS text messaging [107], or computer-delivered interventions [112].

The effects did seem to decline in the long run, and the 2 meta-analyses that explicitly examined effectiveness 1 year postintervention (for internet-based and computer-delivered interventions) found that the effect size declined, although it remained statistically significant [102,109].

Adherence (Considered for All Behavioral Areas)

Typically, adherence data were not reported. Where reported, there were generally decreases in program usage over the intervention period [75]. Rates of attrition were variable and sometimes very high, for instance, in one diet review, it was 0% to 84% [61] and in 2 smoking reviews it was ≥60% [86] and 35% to 84% (median 70%) [93]. Bucking the trend, a couple of reviews reported high adherence to digital physical activity interventions [46,60]. Many reviews found a dose-response relationship, whereby the effectiveness of the intervention was positively associated with dietary usage [20,21,61], weight loss [47], smoking [87,90,93,103], and alcohol [79]. However, there was no unanimity, for instance, one combination review found that the attrition rate was lower in internet-based interventions than in face-to-face settings [103] and 2 others found no evidence that retention influenced outcomes [107,110].

Discussion

Principal Findings

We reviewed 94 systematic reviews and meta-analyses that examined the effectiveness of digital interventions in changing health-related behavior and improving health outcomes in the areas of diet, physical activity, diet and physical activity combined, alcohol consumption, and smoking cessation, alone or in any other combination. The effectiveness of digital interventions differed according to the area of health behavior reviewed. Small positive effects were evident in smoking- and alcohol-related interventions. Similar findings were observed in the combined diet and physical activity interventions, as well as in other outcome combinations. However, there was little evidence of the effectiveness of stand-alone diet interventions, and evidence of the effectiveness of physical activity interventions was mixed, with some consistently positive evidence for mobile interventions and some promising evidence for exergaming. Digital interventions were most effective in the short-to-mid term (approximately 3 to 6 months), but there was insufficient evidence about their long-term effect. Typically, they were more effective than no intervention. There is mixed evidence on their effectiveness compared with nondigital interventions.

Our secondary objective was to identify differences in effectiveness between the modes of delivery of digital interventions. We identified internet-based interventions to be one of the more effective interventions for each area studied, except for physical activity alone. Mobile interventions were particularly effective for diet and physical activity combined (medium effects), but they were also effective for alcohol and smoking (small effects) and physical activity alone. Social media interventions were not effective for diet and physical activity combined (weight loss interventions), they had mixed effects for diet, and there was limited evidence for other areas. Computer-delivered technologies had small effects for smoking and alcohol consumption, but the effects for diet and physical activity were mixed.

The effect sizes reported in the reviews were generally below the National Institute for Health Care Excellence (NICE) guidelines for effectiveness of interventions, although it was often difficult to compare results with NICE guidelines, as different measures were used.

For weight management, NICE guideline PH53 (Recommendation 13) states that commissioned lifestyle weight management programs should have at least a 60% completion rate and should be likely to lead to an average weight loss of at least 3%, with at least 30% of participants losing at least 5% of their weight [114]. It should be noted that this is based on all participants, that is, those who attend at least one session. For those *completing* the service, that is, attending at least 75% of sessions, the key performance indicator of 50% losing at least 5% has been set [115]. In contrast, most reviews reported changes in weight in kilograms or changes in BMI, rather than percentage weight loss. The highest effect sizes for BMI in our review were −0.92 kg/m^2 [49] and −0.43 kg/m^2 [38], which are extremely unlikely to represent a 3% or 5% weight loss in individuals with overweight or obesity. The largest changes in weight found in the reviews of digital interventions were −2.71 kg in one study [72] and −2.56 kg in one meta-analysis [71]. In comparison, the Hartman-Boyce evidence review that supports NICE guideline PH53 found an average effect size of −2.59 kg for face-to-face services at 12 months (intention-to-treat analysis) [116]. Most effect sizes from digital interventions did not reach the effect size of face-to-face services.

For smoking, the national outcome measure for stop smoking services is 4-week quits. Smokers attempting to stop without

XSL•FO
RenderX

additional support generally have a success rate of 25% at 4 weeks for carbon monoxide–validated quits and a success rate of about 35% at 4 weeks for self-reported quits [117]. Therefore, to show an impact, services must achieve success rates equivalent to or in excess of these rates that smokers achieve without support. Patients who receive stop smoking service support (behavioral support and pharmacotherapy) are 3 times more likely to quit than those with no support [118], and there is a cessation rate of 35% for brief intervention services but only at the 4-week point [117]. It is difficult to compare the results of our review with these services because the outcomes of digital smoking interventions are often expressed as ORs for smoking cessation rather than cessation rates. The only available review that used cessation rates demonstrated a cessation rate of 14.8% [92]. This is lower than the observed rate from stop smoking services or brief advice, but the follow-up point was considerably later and cessation rates decreased over time.

For alcohol, the average weekly reduction in drinking from brief advice interventions is 20 g of alcohol (about 2.5 UK units) [119], so the reductions in weekly alcohol consumption achieved by digital alcohol interventions are comparable when including all interventions (22 g of alcohol or 3 UK units) but lower if one restricts attention to high-quality evidence (11 g of alcohol or 1.5 UK units) [77].

To the extent that the effectiveness of digital interventions is below the NICE guidelines, doctors and organizations should be cautious about recommending them to patients who would benefit from behavior change in the 4 areas of our review. However, it may be valuable to recommend them to people who refuse a face-to-face intervention. On the basis of the evidence that we found, digital interventions for weight loss should combine diet and exercise unless they are mobile interventions targeting physical activity.

For digital interventions for both smoking and combined diet and physical activity, there were studies showing significant health effects but not significant behavioral effects. This seems paradoxical. However, it is possible that even a small increase in physical activity or a small improvement in diet, which may not be statistically significant, can improve health markers over an intervention period, particularly in the most inactive and less fit individuals. Light physical activity is beneficial for health outcomes, including cardiometabolic risk factors [120]. We also noted that, although statistically significant, the health effects are small and possibly not meaningful.

Limitations

Owing to the rapid nature of our review, we did not perform full hand searches or consult experts. This may mean that we overlooked some reviews. At the other end of the spectrum, by reviewing reviews, there is the possibility that some studies were double counted if they were covered in more than one review.

Our ability to make inferences from the literature reviewed is limited for the following reasons.

Heterogeneity was consistently high across reviews. Heterogeneity of effects probably reflects heterogeneity of interventions, which could be a consequence of rapid advances in digital devices and systems. There were also heterogeneous outcome measures. As the reviews covered different types of digital interventions and outcome measures, it was difficult to make comparisons. Differing outcome measures may have differentially impacted the effectiveness of modes of intervention or the general effectiveness of interventions in the areas we investigated. For instance, smaller effect sizes were reported in studies addressing smoking use than in studies on alcohol consumption, possibly because studies addressing alcohol use tended to use reductions in drinking behavior as their outcome variable, while studies addressing smoking use tended to apply the more stringent standard of abstinence [112].

Follow-up times were relatively short, so we cannot know if behavior change would be sustained in the long term. Some trials only provided behavioral data, so we cannot be sure of health outcomes. A review of physical activity found that the average rate of sustained use of digital health interventions over 10 weeks was 50% [64]. This is consistent with the findings of another systematic review on physical activity apps, which concluded that apps are effective in the short term (up to 3 months) but not longer [121].

Most trials reported intent-to-treat analysis, and typically, adherence data were not reported. This makes it difficult to assess nonsignificant effects to determine whether they resulted from ineffectiveness of treatment or from nonadherence. Where attrition rates were reported, they were often high.

Anecdotally, digital interventions are being used both to supplement and replace face-to-face services. However, most reviews did not discriminate between these functions. In the domain of weight loss, 5 reviews reported enhanced effects on weight loss in interventions that incorporated personal contact or counseling [51,54,57,62,74]. One meta-analysis showed that infrequent in-person treatment was superior in limiting weight gain to computer-based interventions (mean difference 0.5 kg). Digital interventions that particularly benefit from involving people alongside are thought to include sensitive tailoring of feedback [57,62] and social support [54].

Confidence ratings were critically low in 79 of 93 reviews (85%). However, when isolated, those reviews that were rated critically low presented findings that were consistent with the overall findings: equivocal evidence on effectiveness for diet or physical activity outcomes but consistent findings of short-term effects for alcohol, smoking, and other combined outcomes.

Even when the AMSTAR-2 [16] ratings were moderated (so that justifying any publication language inclusion criterion and providing a list of justifications for excluding reviews were no longer considered critical flaws), 74 of 93 (80%) reviews were rated critically low. During the generation of confidence ratings, it was noted that many reviews failed to satisfy items 2 and 13. These are considered critical items for all review types. Item 2 specifies that, as a minimum, reviews state that a protocol containing research questions, search strategy, inclusion/exclusion criteria, and a risk of bias assessment was completed before conducting the review. Item 13 dictates that reviews should account for the risk of bias in individual studies when interpreting/discussing the results of the review. The

XSL•FO
RenderX

inclusion of these items in the AMSTAR-2 rating may represent an aspiration to improve standards. Our AMSTAR-2 quality ratings are consistent with other evidence that suggests that it is possible to satisfy PRISMA standards yet still have poor methodological quality [122]. However, judging reviews according to such high standards, such that they are virtually all rated as being very low, masks the differences in quality.

Future Research Work

This is the first review of reviews on the effectiveness of digital interventions with such a large scope. It summarizes the state of our knowledge of digital interventions for health improvement behaviors in nonclinical populations. However, the literature could be developed to be more helpful for professionals and organizations who need to decide whether to promote digital interventions, which ones to promote, which areas of behaviors to promote them for, and who to promote them to.

For policy purposes, reviews with mixed controls are of limited use. It matters whether a digital intervention is being compared with no intervention or an active nondigital control. It also matters whether the intervention is a stand-alone digital intervention or whether digital is being used as an adjunct to face-to-face services. We cannot assume that a digital intervention that is successful as an adjunct will also be successful as a stand-alone intervention. Therefore, reviews are needed to separately summarize the evidence base for these different ways of using digital interventions. It could be helpful to have well-structured and coordinated reviews that collate a high-level picture for each area of behavior, which can be updated on a regular basis. We need comparisons with national measures of effectiveness, such as the NICE guidelines, to more easily influence policy.

We also need reviews that can help us determine which interventions are most effective. In the future, it would be helpful to conduct comparative research on the mode of delivery of digital interventions (including comparisons of effect sizes), so that we can determine the most promising interventions for further development. There was also a lack of evidence about the long-term effects of interventions, and more studies on the sustainability of behavior change after digital interventions are needed. It would be especially useful to have this information in comparison with active controls.

Professionals also need to know who should be recommended digital interventions. Therefore, it would also be useful to know whether effect sizes are consistent across various subgroups of the population or whether digital interventions have different effects in different subgroups. We were not specifically looking for this information, but the reviews that were surveyed had mixed findings about whether the effectiveness of digital interventions varied with sex and age. Three meta-analyses found no significant association between sex and effect size or age and effect size [102,109,110]. However, one study found that the effect of interventions declined as age increased [105]. There may also be sex-based differences in adherence, and it is plausible, though not proven, that adherence moderates effect size. One narrative synthesis found that women and middle-aged participants were more likely to use web-based intervention services than men and younger participants, and women were more adherent to the overall intervention [103]. There is even less information about differential effects according to socioeconomic status (SES). In the domain of smoking, one review found that the relative effectiveness of technology-based interventions appeared to be comparable between low- and high-SES groups [83].

The acceptability of digital interventions to their target users also warrants further study. In one review of digital interventions of addictive behaviors, participants expressed a preference for internet-based services because of the convenience and increased confidentiality, and individuals who might not otherwise seek treatment said they would consider an internet-based intervention [103].

Providers may be drawn to digital tools in the hope that they are cost-effective. While not the purpose of our review of reviews, we noted insufficient evidence in the reviews to draw any preliminary inferences about the cost-effectiveness of digital interventions. The evidence from the reviews was mixed. There was evidence in favor (internet-based health interventions [41]), evidence against (adaptive e-learning interventions [22]), and mixed evidence: 1 of 3 (33%) web-based interventions was cost-effective compared with in-person interventions at 6 months [52]. Cost-effectiveness may also depend on whether digital interventions supplement or replace face-to-face interventions. Cost-threshold analyses indicated that some form of electronic intervention is likely to be cost-effective when added to nonelectronic behavioral support, but there is substantial uncertainty with regard to determining the most effective (thus most cost-effective) type of electronic intervention, which warrants further research [84]. Future work will need to investigate cost-effectiveness to allocate resources to developing the most promising digital tools.

Conclusions

Our review of reviews summarizes the state of our knowledge of digital interventions for health improvement behaviors in nonclinical populations. We found positive but small effects for digital interventions that targeted diet and physical activity combined, greater effects—but still small—for smoking and alcohol consumption, and positive, medium-sized effects for mobile interventions for physical activity alone. More high-quality research is needed to assess the sustainability of the effects of digital interventions in the long term, the differences between modes of delivery for digital interventions, their effect on different population subgroups, their cost-effectiveness compared with existing behavior change approaches, and in particular whether they are better used as an adjunct to or replacement for face-to-face treatment. We need the answers to these questions to be able to make an informed decision about whether digital behavior change tools should be integrated into the NHS Health Check program.

XSL•FO
RenderX

Acknowledgments

The authors are grateful to Nicola Pearse-Smith for conducting the searches, Rachel Clarke for advice, and Shevez Jeffers for help with the copyediting. This work was funded by Public Health England (PHE). The authors were employed by, or were interns at, PHE when researching this paper. AY was affiliated with the Centre for Diet and Activity Research, MRC Epidemiology Unit, University of Cambridge, when this work was undertaken and is currently affiliated with the Department of Public Health, Environments & Society, Faculty of Public Health & Policy, London School of Hygiene & Tropical Medicine.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 63 KB - jmir_v23i5e19688_app1.doc]

Multimedia Appendix 2
Search strategy.
[DOCX File , 20 KB - jmir_v23i5e19688_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOC File , 98 KB - jmir_v23i5e19688_app3.doc]

Multimedia Appendix 4
List of excluded reviews.
[XLSX File (Microsoft Excel File), 20 KB - jmir_v23i5e19688_app4.xlsx]

Multimedia Appendix 5
Breakdown of the reviews included in the study.
[DOCX File , 16 KB - jmir_v23i5e19688_app5.docx]

Multimedia Appendix 6
Key characteristics of the included studies.
[XLSX File (Microsoft Excel File), 43 KB - jmir_v23i5e19688_app6.xlsx]

References

1. UK Factsheet. British Heart Foundation. 2021 Mar. URL: https://www.bhf.org.uk/-/media/files/research/heart-statistics/bhf-cvd-statistics-uk-factsheet.pdf [accessed 2021-03-30]
2. The NHS Long Term Plan. National Health Service Long Term Plan. 2019 Jan 07. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2021-03-30]
3. Prevention is better than cure: our vision to help you live well for longer. Department of Health and Social Care. 2018 Jan 23. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/753688/Prevention_is_better_than_cure_5-11.pdf [accessed 2021-03-30]
4. Thomas K. NHS Health Check Best Practice Guidance. Public Health England.: Public Health England; 2016. URL: https://www.networks.nhs.uk/nhs-networks/healthcare-science/chief-scientific-officers-16-17-workstreams/diagnostics-supporting-mental-health-t-f-group/lab-in-a-bag-training-workshops/nhs-health-check-best-practice-guidance [accessed 2021-04-20]
5. Public Health England. NHS Health Check: stocktake and action plan. gov.uk. 2018 Jan 23. URL: https://www.gov.uk/government/publications/nhs-health-check-stocktake-and-action-plan#:~:text=This%20stocktake%20and%20action%20plan,in%20the%20next%205%20years [accessed 2021-03-30]
6. Clifton PM, Keogh JB. Effects of different weight loss approaches on CVD risk. Curr Atheroscler Rep 2018 Apr 25;20(6):1-8. [doi: 10.1007/s11883-018-0728-8]
7. Physical Activity, Health Improvement and Protection. Start Active, Stay Active: A report on physical activity from the four home countries' Chief Medical Officers. Department of Health. 2011 Jul 11. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830943/withdrawn_dh_128210.pdf [accessed 2021-03-30]

XSL•FO
RenderX

8. World Health Organisation. Fact sheet about health benefits of smoking cessation. World Health Organisation. 2019. URL: https://www.who.int/tobacco/quitting/benefits/en/ [accessed 2021-03-30]
9. Holmes M, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, InterAct Consortium. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. Br Med J 2014 Jul 10;349:g4164 [FREE Full text] [doi: 10.1136/bmj.g4164] [Medline: 25011450]
10. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S, et al. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. Br Med J 2016 Mar 15;352:1-11. [doi: 10.1136/bmj.i1102]
11. Shefer G, Silarova B, Usher-Smith J, Griffin S. The response to receiving phenotypic and genetic coronary heart disease risk scores and lifestyle advice – a qualitative study. BMC Public Health 2016 Dec 3;16(1):1-11. [doi: 10.1186/s12889-016-3867-2]
12. Silarova B, Lucas J, Butterworth AS, Di Angelantonio E, Girling C, Lawrence K, et al. Information and Risk Modification Trial (INFORM): design of a randomised controlled trial of communicating different types of information about coronary heart disease risk, alongside lifestyle advice, to achieve change in health-related behaviour. BMC Public Health 2015 Sep 07;15:868 [FREE Full text] [doi: 10.1186/s12889-015-2192-5] [Medline: 26345710]
13. Silarova B, Sharp S, Usher-Smith JA, Lucas J, Payne RA, Shefer G, et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM Randomised Controlled Trial. Heart 2019 Jul;105(13):982-989. [doi: 10.1136/heartjnl-2018-314211] [Medline: 30928969]
14. Murray E, Hekler EB, Andersson G, Collins L, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
16. 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]
17. Carey R. A taxonomy of modes of delivery of behaviour change interventions: development and evaluation. In: The European Health Psychologist. 2017 Dec 31 Presented at: European Health Psychology Society Conference; 2017; Padova, Italy p. 666-666 URL: https://osf.io/d4qyk/
18. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988:1-546.
19. Olivier J, May W, Bell M. Relative effect sizes for measures of risk. Commun Stat Theory Methods 2017 Mar 23;46(14):6774-6781. [doi: 10.1080/03610926.2015.1134575]
20. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
21. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: a systematic literature review. J Telemed Telecare 2015 Jul;21(5):243-253. [doi: 10.1177/1357633X15572203] [Medline: 25680388]
22. Harris J, Felix L, Miners A, Murray E, Michie S, Ferguson E, et al. Adaptive e-learning to improve dietary behaviour: a systematic review and cost-effectiveness analysis. Health Technol Assess 2011 Oct;15(37):1-160 [FREE Full text] [doi: 10.3310/hta15370] [Medline: 22030014]
23. McCarroll R, Eyles H, Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: A systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]
24. Bock C, Jarczok MN, Litaker D. Community-based efforts to promote physical activity: a systematic review of interventions considering mode of delivery, study quality and population subgroups. J Sci Med Sport 2014 May;17(3):276-282. [doi: 10.1016/j.jsams.2013.04.009] [Medline: 23693030]
25. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
26. Bottorff JL, Seaton CL, Johnson ST, Caperchione CM, Oliffe JL, More K, et al. An updated review of interventions that include promotion of physical activity for adult men. Sports Med 2015 Jun;45(6):775-800 [FREE Full text] [doi: 10.1007/s40279-014-0286-3] [Medline: 25430599]
27. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
28. Cheatham S, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]

XSL•FO
RenderX

29. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery W. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
30. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
31. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019 Aug 01;27(4):565-593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
32. Freak-Poli R, Cumpston M, Peeters A, Clemes SA. Workplace pedometer interventions for increasing physical activity. Cochrane Database Syst Rev 2013 Apr 30(4):CD009209. [doi: 10.1002/14651858.CD009209.pub2] [Medline: 23633368]
33. George ES, Kolt GS, Duncan MJ, Caperchione CM, Mummery WK, Vandelanotte C, et al. A review of the effectiveness of physical activity interventions for adult males. Sports Med 2012 Apr 01;42(4):281-300. [doi: 10.2165/11597220-000000000-00000] [Medline: 22350570]
34. Hakala S, Rintala A, Immonen J, Karvanen J, Heinonen A, Sjögren T. Effectiveness of physical activity promoting technology-based distance interventions compared to usual care. Systematic review, meta-analysis and meta-regression. Eur J Phys Rehabil Med 2017 Dec;53(6):953-967. [doi: 10.23736/S1973-9087.17.04585-3] [Medline: 28466628]
35. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
36. Jenkins A, Christensen H, Walker JG, Dear K. The effectiveness of distance interventions for increasing physical activity: a review. Am J Health Promot 2009;24(2):102-117. [doi: 10.4278/ajhp.0801158] [Medline: 19928483]
37. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
38. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
39. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
40. Müller AM, Khoo S. Non-face-to-face physical activity interventions in older adults: a systematic review. Int J Behav Nutr Phys Act 2014 Mar 10;11(1):35 [FREE Full text] [doi: 10.1186/1479-5868-11-35] [Medline: 24612748]
41. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep;22(3):451-469. [doi: 10.1177/1460458214567004] [Medline: 25649783]
42. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
43. Peng W, Lin J, Crouse J. Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):681-688. [doi: 10.1089/cyber.2010.0578] [Medline: 21668370]
44. Song Y. Feasibility and effectiveness of mobile phones in physical activity promotion for adults 50 years and older: a systematic review. Top Geriatr Rehabil 2018;34(3):A-222. [doi: 10.1097/tgr.0000000000000197]
45. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
46. Street TD, Lacey SJ, Langdon RR. Gaming your way to health: a systematic review of exergaming programs to increase health and exercise behaviors in adults. Games Health J 2017 Jun;6(3):136-146. [doi: 10.1089/g4h.2016.0102] [Medline: 28448175]
47. Aalbers T, Baars M, Rikkert MO. Characteristics of effective internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
48. Allen JK, Stephens J, Patel A. Technology-assisted weight management interventions: systematic review of clinical trials. Telemed J E Health 2014 Dec;20(12):1103-1120 [FREE Full text] [doi: 10.1089/tmj.2014.0030] [Medline: 25409001]
49. An R, Ji M, Zhang S. Effectiveness of social media-based interventions on weight-related behaviors and body weight status: review and meta-analysis. Am J Health Behav 2017 Nov 01;41(6):670-682. [doi: 10.5993/AJHB.41.6.1] [Medline: 29025495]
50. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]

XSL•FO
RenderX

51. Bassi N, Karagodin I, Wang S, Vassallo P, Priyanath A, Massaro E, et al. Lifestyle modification for metabolic syndrome: a systematic review. Am J Med 2014 Dec;127(12):1242.e1-1242.e10. [doi: 10.1016/j.amjmed.2014.06.035] [Medline: 25004456]
52. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, De Fátima H Diniz M, Ribeiro AL. Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
53. Carvalho de Menezes M, Bedeschi LB, Santos LC, Lopes AC. Interventions directed at eating habits and physical activity using the Transtheoretical Model: a systematic review. Nutr Hosp 2016 Sep 20;33(5):586. [doi: 10.20960/nh.586] [Medline: 27759990]
54. Coons MJ, Demott A, Buscemi J, Duncan JM, Pellegrini CA, Steglitz J, et al. Technology interventions to curb obesity: a systematic review of the current literature. Curr Cardiovasc Risk Rep 2012 Apr;6(2):120-134 [FREE Full text] [doi: 10.1007/s12170-012-0222-8] [Medline: 23082235]
55. Dutton GR, Laitner MH, Perri MG. Lifestyle interventions for cardiovascular disease risk reduction: a systematic review of the effects of diet composition, food provision, and treatment modality on weight loss. Curr Atheroscler Rep 2014 Oct;16(10):442 [FREE Full text] [doi: 10.1007/s11883-014-0442-0] [Medline: 25092578]
56. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
57. 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]
58. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
59. Liu F, Kong X, Cao J, Chen S, Li C, Huang J, et al. Mobile phone intervention and weight loss among overweight and obese adults: a meta-analysis of randomized controlled trials. Am J Epidemiol 2015 Mar 01;181(5):337-348 [FREE Full text] [doi: 10.1093/aje/kwu260] [Medline: 25673817]
60. Lyzwinski L. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Personalized Med 2014 Jun 30;4(3):311-385. [doi: 10.3390/jpm4030311] [Medline: 25563356]
61. 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 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
62. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011 Mar 04;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
63. Maon S, Edirippulige S, Ware R, Batch J. The use of web-based interventions to prevent excessive weight gain. J Telemed Telecare 2012 Jan;18(1):37-41. [doi: 10.1258/jtt.2011.110306] [Medline: 22101608]
64. Maxwell K. Technology as a Health Intervention and the Self-Efficacy of Men. Walden University. 2015. URL: https://scholarworks.waldenu.edu/dissertations/418/ [accessed 2021-04-20]
65. Mita G, Ni Mhurchu C, Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2016 Apr;74(4):237-247 [FREE Full text] [doi: 10.1093/nutrit/nuv106] [Medline: 26946250]
66. Podina IR, Fodor LA. Critical review and meta-analysis of multicomponent behavioral e-health interventions for weight loss. Health Psychol 2018 Jun;37(6):501-515. [doi: 10.1037/hea0000623] [Medline: 29733617]
67. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
68. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
69. Seo D, Niu J. Evaluation of internet-based interventions on waist circumference reduction: a meta-analysis. J Med Internet Res 2015 Jul 21;17(7):e181 [FREE Full text] [doi: 10.2196/jmir.3921] [Medline: 26199208]
70. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
71. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
72. Tang JC, Abraham C, Greaves CJ, Nikolaou V. Self-directed interventions to promote weight loss: a systematic review and meta-analysis. Health Psychol Rev 2016 May 10;10(3):358-372. [doi: 10.1080/17437199.2016.1172979]
73. Tsoli S, Sutton S, Kassavou A. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. BMJ Open 2018 Feb 24;8(2):e018974. [doi: 10.1136/bmjopen-2017-018974]

XSL•FO
RenderX

74. Wieland L, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012 Aug 15(8):CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
75. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014 Feb 12;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
76. Willis EA, Szabo-Reed AN, Ptomey LT, Steger FL, Honas JJ, Washburn RA, et al. Do weight management interventions delivered by online social networks effectively improve body weight, body composition, and chronic disease risk factors? A systematic review. J Telemed Telecare 2017 Feb;23(2):263-272 [FREE Full text] [doi: 10.1177/1357633X16630846] [Medline: 26880695]
77. Kaner E, 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 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
78. Kolar C, von Treuer K. Alcohol misuse interventions in the workplace: a systematic review of workplace and sports management alcohol interventions. Int J Ment Health Addiction 2015 Jun 20;13(5):563-583. [doi: 10.1007/s11469-015-9558-x]
79. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun 30;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
80. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
81. Vernon ML. A review of computer-based alcohol problem services designed for the general public. J Subst Abuse Treat 2010 Apr;38(3):203-211 [FREE Full text] [doi: 10.1016/j.jsat.2009.11.001] [Medline: 20015607]
82. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016 Sep;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
83. Boland V, Stockings EA, Mattick RP, McRobbie H, Brown J, Courtney RJ. The methodological quality and effectiveness of technology-based smoking cessation interventions for disadvantaged groups: a systematic review and meta-analysis. Nicotine Tob Res 2018 Feb 07;20(3):276-285. [doi: 10.1093/ntr/ntw391] [Medline: 28034998]
84. Chen Y, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012 Oct;16(38):1-205, iii [FREE Full text] [doi: 10.3310/hta16380] [Medline: 23046909]
85. Cheung KL, Wijnen B, de Vries H. A review of the theoretical basis, effects, and cost effectiveness of online smoking cessation interventions in the netherlands: a mixed-methods approach. J Med Internet Res 2017 Jun 23;19(6):e230 [FREE Full text] [doi: 10.2196/jmir.7209] [Medline: 28645889]
86. Danielsson A, Eriksson A, Allebeck P. Technology-based support via telephone or web: a systematic review of the effects on smoking, alcohol use and gambling. Addict Behav 2014 Dec;39(12):1846-1868. [doi: 10.1016/j.addbeh.2014.06.007] [Medline: 25128637]
87. Gainsbury S, Blaszczynski A. A systematic review of Internet-based therapy for the treatment of addictions. Clin Psychol Rev 2011 Apr;31(3):490-498. [doi: 10.1016/j.cpr.2010.11.007] [Medline: 21146272]
88. Graham A, Carpenter K, Cha S, Cole S, Jacobs M, Raskob M, et al. Systematic review and meta-analysis of internet interventions for smoking cessation among adults. Subst Abuse Rehabil 2016;7:55-69 [FREE Full text] [doi: 10.2147/SAR.S101660] [Medline: 27274333]
89. Health Information and Quality Authority. Health technology assessment (HTA) of smoking cessation interventions. Health Information and Quality Authority. 2017. URL: https://www.hiqa.ie/sites/default/files/2017-04/Smoking%20Cessation%20HTA.pdf [accessed 2021-03-30]
90. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]
91. McCrabb S, Baker AL, Attia J, Skelton E, Twyman L, Palazzi K, et al. Internet-based programs incorporating behavior change techniques are associated with increased smoking cessation in the general population: a systematic review and meta-analysis. Ann Behav Med 2019 Feb 01;53(2):180-195. [doi: 10.1093/abm/kay026] [Medline: 29750240]
92. 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]
93. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]

XSL•FO
RenderX

94. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
95. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
96. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
97. Taylor G, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
98. 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:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
99. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information technology and lifestyle: a systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. J Am Heart Assoc 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
100. Aneni EC, Roberson LL, Maziak W, Agatston AS, Feldman T, Rouseff M, et al. PLoS One 2014;9(1):e83594 [FREE Full text] [doi: 10.1371/journal.pone.0083594] [Medline: 24421894]
101. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
102. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
103. Chebli J, Blaszczynski A, Gainsbury SM. Internet-based interventions for addictive behaviours: a systematic review. J Gambl Stud 2016 Dec;32(4):1279-1204. [doi: 10.1007/s10899-016-9599-5] [Medline: 27002522]
104. Covolo L, Ceretti E, Moneda M, Castaldi S, Gelatti U. Does evidence support the use of mobile phone apps as a driver for promoting healthy lifestyles from a public health perspective? A systematic review of Randomized Control Trials. Patient Educ Couns 2017 Dec;100(12):2231-2243. [doi: 10.1016/j.pec.2017.07.032] [Medline: 28855063]
105. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
106. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
107. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
108. Hou S, Charlery SR, Roberson K. Systematic literature review of Internet interventions across health behaviors. Health Psychol Behav Med 2014 Jan 01;2(1):455-481 [FREE Full text] [doi: 10.1080/21642850.2014.895368] [Medline: 25750795]
109. 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]
110. Lustria ML, 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]
111. Palmer M, Sutherland J, Barnard S, Wynne A, Rezel E, Doel A, et al. The effectiveness of smoking cessation, physical activity/diet and alcohol reduction interventions delivered by mobile phones for the prevention of non-communicable diseases: a systematic review of randomised controlled trials. PLoS One 2018;13(1):e0189801 [FREE Full text] [doi: 10.1371/journal.pone.0189801] [Medline: 29304148]
112. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]
113. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
114. Weight management: lifestyle services for overweight or obese adults. National Institute for Health and Care Excellence. 2014. URL: https://www.nice.org.uk/guidance/ph53/resources/weight-management-lifestyle-services-for-overweight-or-obese-adults-pdf-1996416726469 [accessed 2021-03-30]
115. Thomson L. Key Performance Indicators: Tier 2 Weight Management Services for Adults. Public Health England. 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/656531/adult_weight_management_key_performance_indicators.pdf [accessed 2021-03-30]

XSL•FO
RenderX

116. Hartmann-Boyce J, Aveyard P, Koshiaris C, Jebb SA. Development of tools to study personal weight control strategies: OxFAB taxonomy. Obesity (Silver Spring) 2016 Feb;24(2):314-320 [FREE Full text] [doi: 10.1002/oby.21341] [Medline: 26748902]
117. West R, May S, West M, Croghan E, McEwen A. Performance of English stop smoking services in first 10 years: analysis of service monitoring data. Br Med J 2013 Aug 19;347:f4921 [FREE Full text] [doi: 10.1136/bmj.f4921] [Medline: 23963106]
118. West R, Papadakis S. Stop smoking services: increased chances of quitting. National Centre for Smoking Cessation and Training. 2019. URL: https://www.ncsct.co.uk/usr/pub/Stop%20smoking%20services%20effectiveness.pdf [accessed 2021-03-30]
119. Kaner E, Beyer F, Dickinson HO, Pienaar E, Campbell F, Schlesinger C, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2007 Apr 18(2):CD004148. [doi: 10.1002/14651858.CD004148.pub3] [Medline: 17443541]
120. Amagasa S, Machida M, Fukushima N, Kikuchi H, Takamiya T, Odagiri Y, et al. Is objectively measured light-intensity physical activity associated with health outcomes after adjustment for moderate-to-vigorous physical activity in adults? A systematic review. Int J Behav Nutr Phys Act 2018 Jul 09;15(1):65 [FREE Full text] [doi: 10.1186/s12966-018-0695-z] [Medline: 29986718]
121. 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]
122. Johnson BT, Hennessy EA. Systematic reviews and meta-analyses in the health sciences: best practice methods for research syntheses. Soc Sci Med 2019 Jul;233:237-251. [doi: 10.1016/j.socscimed.2019.05.035] [Medline: 31233957]

Abbreviations

CVD: cardiovascular disease
NHS: National Health Service
NICE: National Institute for Health Care Excellence
PHE: Public Health England
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio
SES: socioeconomic status
SMD: standardized mean difference

Edited by R Kukafka; submitted 04.06.20; peer-reviewed by S Hakala, T Stütz; comments to author 14.07.20; revised version received 05.10.20; accepted 03.02.21; published 14.05.21.

Please cite as:
Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T
Effectiveness of Digital Interventions for Reducing Behavioral Risks of Cardiovascular Disease in Nonclinical Adult Populations: Systematic Review of Reviews
J Med Internet Res 2021;23(5):e19688
URL: https://www.jmir.org/2021/5/e19688
doi:10.2196/19688
PMID:33988126

©Natalie Gold, Amy Yau, Benjamin Rigby, Chris Dyke, Elizabeth Alice Remfry, Tim Chadborn. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.05.2021. 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

Maximizing Participant Engagement, Participation, and Retention in Cohort Studies Using Digital Methods: Rapid Review to Inform the Next Generation of Very Large Birth Cohorts

Joanna Nkyekyer[1,2], BSc, MSc, PhD; Susan A Clifford[1,2], PhD; Fiona K Mensah[1,2], PhD; Yichao Wang[1,2], PhD; Lauren Chiu[3], MD; Melissa Wake[1,2], MD

[1]Murdoch Children's Research Institute, Parkville, Australia

[2]Department of Paediatrics, The University of Melbourne, Parkville, Australia

[3]Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Australia

Corresponding Author:
Melissa Wake, MD
Murdoch Children's Research Institute
50 Flemington Road
Parkville, VIC 3052
Australia
Phone: 61 3 9345 5937
Email: melissa.wake@mcri.edu.au

Abstract

Background: Many current research needs can only be addressed using very large cohorts. In such studies, traditional one-on-one phone, face-to-face, or paper-based engagement may not be feasible. The only realistic mechanism for maintaining engagement and participation at this scale is via digital methods. Given the substantial investment being made into very large birth cohort studies, evidence for optimal methods of participant engagement, participation, and retention over sustained periods without in-person contact from researchers is paramount.

Objective: This study aims to provide an overview of systematic reviews and meta-analyses evaluating alternative strategies for maximizing participant engagement and retention rates in large-scale studies using digital methods.

Methods: We used a rapid review method by searching PubMed and Ovid MEDLINE databases from January 2012 to December 2019. Studies evaluating at least 1 e-engagement, participation, or retention strategy were eligible. Articles were screened for relevance based on preset inclusion and exclusion criteria. The methodological quality of the included reviews was assessed using the AMSTAR-2 (Assessing the Methodological Quality of Systematic Reviews 2) measurement tool, and a narrative synthesis of the data was conducted.

Results: The literature search yielded 19 eligible reviews. Overall, 63% (n=12) of these reviews reported on the effectiveness of e-engagement or participation promotion strategies. These evaluations were generally not conducted within very large observational digital cohorts. Most of the contributing reviews included multipurpose cohort studies (with both observational and interventional elements) conducted in clinical and research settings. Email or SMS text message reminders, SMS text messages or voice notifications, and incentives were the most commonly used design features to engage and retain participants. For parental outcomes, engagement-facilitation interventions influenced uptake and behavior change, including video feedback, goal setting, and intensive human facilitation and support. Participant-stated preferences for content included new knowledge, reminders, solutions, and suggestions about health issues presented in a clear, short, and personalized way. Perinatal and postpartum women valued self-monitoring and personalized feedback. Digital reminders and multiple SMS text messages were specific strategies that were found to increase adherence to medication and clinic attendance, respectively.

Conclusions: This review adds to the growing literature evaluating methods to optimize engagement and participation that may apply to large-scale studies using digital methods; it is promising that most e-engagement and participation promotion strategies appear to be effective. However, these reviews canvassed relatively few strategies, suggesting that few alternative strategies have been experimentally evaluated. The reviews also revealed a dearth of experimental evidence generated within very large observational digital cohort studies, which may reflect the small number of such studies worldwide. Thus, very large studies may need to proactively build in experimental opportunities to test engagement and retention approaches to enhance the success of their own and other large digital contact studies.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e23499) doi:10.2196/23499

KEYWORDS

cohort studies; communication modes; digital study; mobile phone; participant engagement; research methodology; retention; systematic reviews

Introduction

Background

Adult cohort studies (such as the UK Biobank, recruiting 500,000 participants and costing approximately £250 million (US $349 million) to date [1]) have demonstrated the power of mega-cohorts to transform the speed, precision, and capacity for high-value new knowledge for health and health care delivery. Unfortunately, high-profile *early life* initiatives of similar size and ambition, such as the US National Children's Study and UK Life Study, were withdrawn despite £0.8 billion (US $1.2 billion) and £38 million (US $59 million) funding, respectively [2,3], in large part because they stumbled at the first hurdle of engagement and uptake. Others, though successful in recruitment, have had substantial attrition over time [4]. Thus, a limited science of engagement and retention poses a critical hurdle to such studies in meeting their vision of advancing human health.

Engagement is defined as "the extent to and manner in which people actively use a resource and has been operationalized as a multistage process involving the point of engagement, a period of sustained engagement, disengagement, and reengagement" [5]. Many factors may influence the engagement process at different time points. In a research study, indicators of poor engagement may include low initial uptake from the first point of contact or reduced interaction over time, in some cases leading to complete disengagement or dropout. Engagement strategies have been developed to enable cohort studies—both observational and interventional—to meet their aims (eg, improving health behaviors and outcomes) by allowing regular, sustainable engagement with large numbers of participants via remote or digital-only studies [6].

e-Engagement incorporates the participation, recruitment, and retention of participants through digital platforms. Factors that may improve participant e-engagement include its technical features, content, frequency of waves, and engagement-facilitation interventions (EFIs) [7]. User characteristics and digital platform features should also be considered. Ritterband et al [8] simplified this in their internet intervention model, hypothesizing that behavior change is influenced by the stepwise progression of environmental factors, support, and website characteristics affecting adherence, which then affect behavior change (ie, sustained participant engagement) through various mechanisms of change. Thus, maximizing e-engagement can improve the efficiency of research processes and reduce both administration costs [9] and the validity and power costs of significant and systematic nonuptake and attrition [10] in major studies.

Given the expense of longitudinal cohort studies, effective strategies that engage and retain cohort participants are critical to the integrity of research outcomes [11,12]. The retention of study participants is vital to ensure the power and internal validity of longitudinal research [13-15], whereas participant engagement is important for evaluating the efficacy and generalizability of the program under study. A review of randomized controlled trials [16] suggests that delays in participant recruitment or high dropout rates postrandomization may lead to uncertainty in treatment effectiveness and possibly confound results. For example, in the case of technology-based intervention studies, the technology may change over time if recruitment is prolonged, potentially leading to artifacts or differential effects on treatment outcomes. Proposed retention strategies involving (1) contact and scheduling methods, (2) visit characteristics, (3) study personnel, (4) nonfinancial incentives, (5) financial incentives, (6) reminders, (7) special tracking methods, (8) study description, (9) benefits of study, (10) reimbursement, (11) study identity, and (12) community involvement [17,18] may influence participant retention rates. However, there is limited experimental evidence and data for the in-depth exploration of retention strategies and their implementation.

With the recent growth in experimental research on the optimization of digital methods in longitudinal cohort research studies, there is a need for the literature to be collectively synthesized at a pace reflecting the rapid evolution of technology.

Objective

The objective of this review is to provide an overview of strategies that enhance engagement, participation, and retention rates in large-scale digital contact studies, comparing digital methods with alternative (digital and nondigital) methods. This work has been undertaken as part of the design of the forthcoming Generation Victoria (GenV) study [19]. GenV is a whole-of-state birth and parent cohort being planned in the state of Victoria, Australia. After initial face-to-face recruitment, the majority of contact with study participants will be via digital methods. The findings of this review are expected to inform GenV and other very large birth cohort studies in planning.

Methods

Protocol Registration

The protocol of this rapid review was registered with PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42020155430). We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement to report our systematic review [20].

Research Questions and Definitions

In the context of the administration of large-scale digital contact cohort studies, we investigated the following research questions:

XSL•FO
RenderX

1. What technical design features aid engagement, participation, and retention?
2. What EFIs aid engagement, participation, and retention?
3. What feedback is valued by parents with young children?
4. How effective are e-engagement, participation, and retention interventions?

We used the following definitions throughout:

- Engagement: the proportion of participants who receive, open, and actively engage in a survey or an assessment wave. Incorporates the study being able to contact the participant, and the participant being motivated to start the activity.
- EFI: the approach used to increase the acceptability of a web-based program.
- Participation: the proportion of participants who completed a survey or an assessment. Incorporates the participant having time to complete the activities, understanding how to complete the activities, and being willing to provide information about themselves and their family.
- Retention: the proportion of participants who participate across successive waves. Incorporates the study being able to contact the participant and the participant wanting to continue to participate.
- Review: reporting on overall findings of an included systematic review.
- Study: reporting on findings of an individual study reported within a systematic review.

Electronic Searches

Four authors (MW, JN, SAC, and YW) developed the search strategy and refined the searches with an experienced librarian. The search queries used to retrieve our systematic reviews and meta-analyses are presented in Multimedia Appendix 1. Literature searches were performed in PubMed and Ovid MEDLINE databases using both MeSH and free-text words. The results from each search engine were downloaded into an EndNote (Clarivate Analytics) reference library and saved in Covidence (Veritas Health Innovation Ltd). Duplicate studies across the combined groups were removed. We also consulted experts and manually searched for relevant studies.

Selection of Reviews

Two authors (selected from JN, YW, and LC) independently screened each paper title and abstract for relevance. The full text of the remaining papers was independently screened by two of the authors (selected from JN, YW, SAC, and LC). Any disagreements were resolved by consensus. To ensure a standardized process for our review, the author's pilot-tested titles and abstracts, and full-text screening with a sample of papers. This information helped refine the inclusion and exclusion criteria.

Inclusion Criteria

Study Types

As this was a rapid review, we examined only systematic reviews and meta-analyses (not individual source studies) pertinent to large-scale cohort studies. We included studies with observational and/or interventional elements conducted in clinical and research settings. As digital technology is moving so rapidly, we limited our search to reviews published between January 2012 and December 2019, reasoning that these would include relevant older studies while being most technologically relevant to the needs of cohorts being planned in the 2020s.

Studies were eligible if they evaluated at least one of the following e-engagement, participation, or retention strategies (note that testing these strategies could occur in the context of a trial of therapeutic intervention):

1. Alternative contact metrics: for example, frequency per month or year; time of the month, week, or day; duration of each contact; and reminder content and frequency.
2. Reimbursement and gifts or penalties: for example, payments for survey completion, small gifts, or store discount codes.
3. Feedback features: for example, presented as participant's responses or performance at the point of completion; progress over time (with or without comparison with the population); a report sharable with care providers; and thank you certificates.
4. Content features: for example, assessments relevant to the life course approach or development stage of participants and/or their child; the balance of positive and negatively framed questions; ease of understanding; cognitive burden of assessment or survey items; and interest.
5. Technical and design features: for example, native or web app, can leave and return to assessment, the appearance of the interface, gamified interface, and visual progress tracker.
6. Study design features: for example, messages personalized with participant names and study staff contactable to answer questions.
7. Target participant characteristics: for example, demographic, motivation, or burden of disease.
8. Communication modes: for example, visual, auditory, text, and real person or avatar.

Participants

As the respondents in large birth cohorts are usually parents for the first decade of life, our primary focus was adults aged <50 years. Where evidence existed, we considered parents of children aged between 0 and 5 years.

Comparators

Alternative *standard* delivery strategies such as mail, fax, and other digital interventions (DIs).

Outcome Measures

Participant engagement in, completion of, and retention in digital study (survey and assessment) waves throughout short and long periods.

Exclusion Criteria

For initial title and abstract screening, we excluded the following publications: the primary focus (participant) was adults ≥50 years or children as the primary respondents; publications not written in English; and publications with full text not accessible through the University of Melbourne library.

XSL•FO
RenderX

Additional criteria for the full-text review screening were not reporting our outcome metrics of engagement, participation, or retention; focusing on low-income countries; and focusing on rare or uncommon conditions such as HIV or cancer.

Data Extraction

A data extraction template was developed and piloted by the authors (JN, YW, SAC, and LC) in 3 reviews. The template contained general review information (author and search dates), characteristics of included studies (number of relevant studies, study designs, health topics, population age, and geographic area), e-engagement, participation, and retention promoting strategies, the methodological quality of systematic reviews, and a summary of review results and conclusions.

Data were extracted independently by 2 of the authors (selected from JN, YW, SAC, and LC), and any discrepancies were resolved by consensus.

Data Synthesis

A meta-analysis was not performed because of the heterogeneity of intervention types, study designs, study populations, and outcome variables among the included studies. Instead, a narrative summary of the findings across studies was created based on study outcomes (ie, participant engagement, participation, completion, and retention) and strategies promoting these study outcomes.

Methodological Quality of Included Reviews

Two authors (JN and YW) independently assessed the quality of the included review methodology using AMSTAR-2 (Assessing the Methodological Quality of Systematic Reviews 2; a measurement tool to assess systematic reviews) [21]. The appraisal tool of AMSTAR-2 included 16 domains: whether there was a description of the PICO (population, intervention, control group, and outcome) components in the research questions and the inclusion criteria; the protocol of systematic review or meta-analysis; study design rationale; the literature searching strategy; study selection; data extraction; specific details of inclusion and exclusion criteria; adequate detail of the included studies; bias risk assessment of the included studies; the funding sources; appropriate statistical methods; the impact evaluation of the individual study's risk of bias (RoB); the explanation of RoB in individual studies; a satisfactory explanation for any heterogeneity; adequate investigation of publication bias; and potential conflicts of interest. The answer options for the AMSTAR-2 were yes, partial yes, and no. *Yes* denoted a positive result. *No* represented that there was not enough information about the domain. *Partial yes* represented that it partially adhered to the standard. Discrepancies were resolved by consensus.

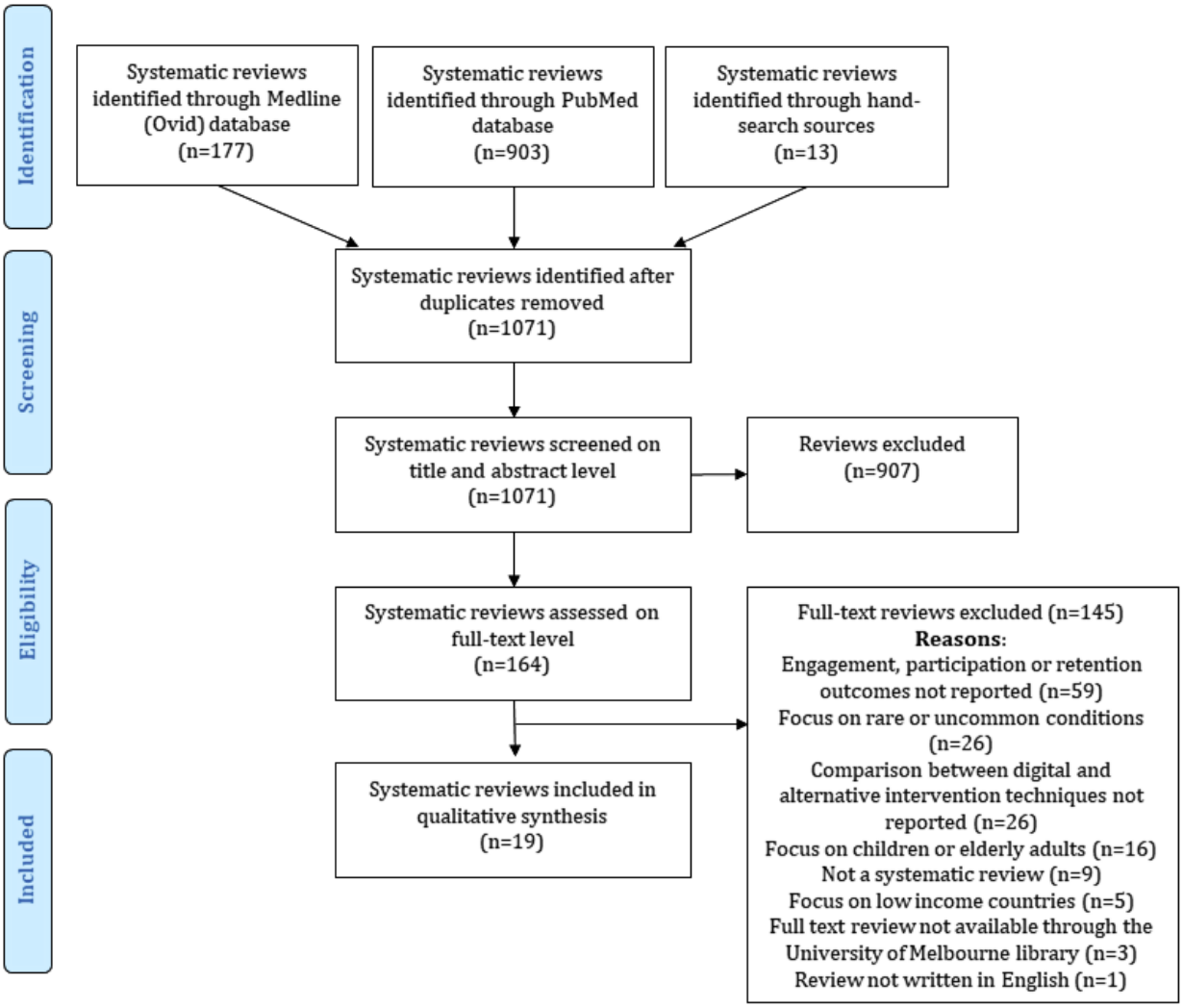
Results

Search and Screening Results

The search strategy yielded 1080 systematic reviews. An additional 13 reviews were identified through a manual search of publications' reference lists. Following the removal of duplicates, 1071 publications were screened. The title and abstract screening excluded 907 reviews. A total of 164 articles underwent full-text review, of which 19 publications [22-40] met the inclusion criteria. Figure 1 summarizes the search and screening process presented in the PRISMA format.

XSL•FO
RenderX

Figure 1. Search and screening process.



Characteristics of Included Reviews

On simple summing, the 19 reviews contained 437 studies and more than 556,000 participants (4 studies did not report the number of participants). Some studies and therefore participants are included in more than one review. Given that we aimed to identify strategies that may influence the outcomes of interest rather than synthesize an overall estimate (as per meta-analytic techniques), we did not see an overlap as problematic.

The characteristics of the included reviews are summarized in Table 1. Most reviews contained studies of varying designs, spanning quantitative and qualitative analyses.

The 19 systematic reviews included studies conducted in both research (8 reviews) and clinical or health care (11 reviews) settings. Of these reviews, 4 examined young people and adults with mental disorders, anxiety, and depressive symptoms. Others examined engagement, participation, and/or retention in digital contact studies among perinatal and postpartum women, patients with chronic diseases, parents in neonatal intensive care units, human papillomavirus vaccine uptake in young children, vaccinations in adults, pregnant women, and uptake of preschool vaccinations.

XSL•FO
RenderX

Table 1. Systematic review characteristics.

Study	Objective	Studies in review	Participants	Population	Regions	Designs of included studies
Research setting						
Alkhaldi, 2016 [23]	Evaluate the effectiveness of tech-based prompts (eg, SMS text messages or calls) for promoting engagement with digital interventions	14	8774	Adults participating in digital interventions for physical and/or mental health	Europe, United States, and Australia	RCTs[a]
Baumeister, 2014 [26]	Investigate the impacts of guidance (human support) on the effectiveness of web-based mental health interventions	14	Not stated	Adults with clinical or subthreshold mental disorders	Not stated	RCTs
Lattie, 2019 [32]	Evaluate factors associated with the effectiveness of web-based mental health interventions	89	15,857	Postsecondary (eg, university) students targeted by universal prevention or treatment intervention programs	Europe, United States, Canada, China, Mexico, Australia, and New Zealand	RCTs, nonrandomized studies, and qualitative
Lim, 2019 [33]	Explore postpartum women's and health professionals' perspectives of digital health interventions for lifestyle management in postpartum women	9	484	Postpartum women	United Kingdom, United States, Bangladesh, and Australia	Qualitative (focus groups or interviews), questionnaire, and observational
Thakkar, 2016 [37]	Investigate the effect of SMS text messaging on medication adherence in chronic disease	16	2742	Adults with chronic diseases, including HIV infection, cardiovascular disease, asthma, diabetes, and epilepsy	Europe, South America, United States, Asia, and Africa	RCTs
Tromp, 2015 [38]	Investigate motivations of children and their parents to participate in clinical drug research	42	Not stated	Children with health conditions (eg, cancer, respiratory diseases, or diabetes) or no health conditions and their parents or guardians	Not stated	Quantitative (questionnaires or registries) and qualitative (interviews, focus groups, or case study)
Valimaki, 2017 [39]	Summarize the content and effectiveness of web-based interventions for depression and anxiety	27	7786	Young people (aged 10-24 years) with symptoms and/or diagnosis of depression or anxiety	Europe, United States, Canada, China, Australia, and New Zealand	RCTs
Whitaker, 2017 [40]	Describe the extent and effectiveness of using Facebook to recruit participants for health research	35	Median 264 per study	People (aged ≥13 years) targeted for recruitment into health studies and interventions, most commonly smoking cessation, human papillomavirus vaccination, and healthier lifestyle interventions	Germany, United States, Canada, Japan, and Australia	Quantitative and qualitative
Clinical or health care setting						
Adams, 2015 [22]	Provide evidence on the effectiveness, acceptability, economic costs, and consequences of parental financial incentives and quasi-mandatory schemes for increasing the uptake of preschool vaccinations	11	334,476	Parent of preschool children living in high-income countries; member of any relevant stakeholder group living in high-income countries	United Kingdom	RCTs and controlled pre-post and time-series analyses

XSL•FO
RenderX

Study	Objective	Studies in review	Participants	Population	Regions	Designs of included studies
Ames, 2019[24]	Describe clients' experiences of receiving health information via their mobile phone	35	Not stated	Adolescent and adult clients of pregnancy, newborn, and child health, sexual health, and family planning health services receiving communication via their mobile devices	United Kingdom, United States, Canada, Southeast Asia, Australia, South America, and Africa	Qualitative study
Atkinson, 2019 [25]	Evaluate the effectiveness of digital push interventions in improving vaccine uptake and series completion compared with nondigital interventions	13	24,224	Adults receiving vaccines themselves, including pregnant women, or parents of adolescents and children eligible for vaccination	United States, Lebanon, Zimbabwe, and Guatemala	RCTs
Belisario, 2015 [27]	Compare the quality of survey responses collected using mobile apps vs other methods	14	2272	Smartphone and tablet apps as a delivery mode in clinical patients. Data collected from participants completing health-related, self-administered questionnaires	Western Europe, United States, Canada, and Korea	RCTs, crossover, and paired repeated measures studies
Dol, 2017 [28]	Examine the effect of eHealth interventions used in neonatal intensive care units on parents and infants	8	Not stated	Parents in neonatal intensive care units	United States, Singapore, the Netherlands, South Korea, and Israel	RCTs, quasi-experimental, pre-post, observational studies, descriptive studies, and prospective studies
Dubad, 2017 [29]	Evaluate the efficacy and usability of mobile mood-monitoring apps in young people	25	110,051	Healthy participants, participants from clinical populations, including youth with a range of mental health, emotional, or behavioral problems	Western Europe, United States, and Australia	RCTs, secondary analyses, nonexperimental studies, and quasi-experimental
Garrido, 2019 [30]	Examine the effectiveness of digital mental health interventions for depression and anxiety in young people	41	16,874	Young people with depression and anxiety	Northern Europe, United States, South America, Asia, and Australia	RCTs, single cohort (including pre-post design), and case studies
Kang, 2017 [31]	Evaluate the impacts of digital interventions on human papillomavirus vaccination	5	14,107	Young adults (males and females) who had received their first human papillomavirus vaccine dose	United States	RCTs
Mertens, 2019 [34]	Evaluate the effects of technology-supported lifestyle interventions on gestational weight gain and postnatal weight loss	9	2603	Perinatal women during pregnancy or within the first postnatal area	United States, Australia, and Iran	RCTs
Parsons et al, 2017 [35]	Evaluate the remotely delivered interventions for children with autism spectrum disorder living outside of urban areas: systematic review	9	197	Families having a child with autism spectrum disorder, living outside of urban areas, and having limited access to services	United States, Canada, and Australia	Pre-post, multiple-base design, RCTs, and quasi-experimental studies
Robotham, 2016 [36]	Assess the impact of digital notifications to improve attendance in clinics	21	16,076	Patients attending health care services	Europe, United States, Asia, Africa, and Australia	RCTs

[a]RCT: randomized controlled trial.

Research Question 1: Technical Design Features That Aid Engagement, Participation, and Retention

We found 4 reviews [22,25,31,32] reporting on several technical design features to aid engagement, participation, and retention, including financial incentives (including gifts), digital pushes (SMS text message alerts), voice notifications, and email or SMS text message reminders and studies' technical feasibility and usability (ie, informed consent). Email or SMS text message reminders and SMS text message notifications were reported

XSL•FO
RenderX

in 4 reviews [31,32,36,37] as the most commonly used technical design feature to improve participation and completion rates. Two reviews reported the use of email or SMS text message reminders and voice notifications [36,37] to enhance study participation. Robotham et al [36] compared zero, one, and multiple SMS text message notifications and voice notifications.

Research Question 2: EFIs That Aid Engagement, Participation, and Retention

The following EFIs were reported by 5 reviews [24,26,28,30,35] as a means to aid uptake and participant engagement.

SMS Text Messages and Interactive Voice Response Messages

The review by Ames et al [24] examined perceptions and experiences of digital targeted client communication (ie, SMS text messages and interactive voice response messages) via mobile devices in the areas of reproductive, maternal, newborn, and adolescent health. The results suggested that many clients liked receiving messages from health services using mobile phones. Content preferences included new knowledge, reminders, solutions, and suggestions about health issues presented in a clear, short, and personalized way.

Intensive Guidance (Web-Based Interventions With Human Facilitation, Support, or Coaches)

According to the review by Baumeister et al [26], in treating mental health disorders, guidance, as a retention strategy, improved rates of completion (pooled completer rate: odds ratio 2.76, 95% CI 1.68-4.53; n=6) and the number of completed modules (pooled mean number of completed modules: standardized mean difference 0.52, 95% CI 0.37-0.067; n=7). Lim et al [33] reported on the use of lifestyle coaching as an EFI to aid DI uptake. DIs were perceived as positive, user-friendly, and acceptable. Engagement strategies employed in DIs were monitoring and feedback, goal setting, health professional input, and social support.

Videoconferencing and Video-Feedback Interventions

The review by Dol et al [28] reported the following EFIs to aid uptake across included studies: Baby CareLink (an educational and emotional support system for parents with children in the neonatal intensive care unit) [41], Skype, and FaceTime. In this review, no significant differences were found between parents who participated in an e-intervention or received standard care in terms of their reported anxiety and/or stress, possibly because of the greatly varied study design and type of eHealth technology across studies.

Garrido et al [30] summarized the use of web-based modules, learning materials, or activities; group chats or courses; online forums; web-based chat facilities with a mental health professional; games; and psychoeducational computer programs as EFIs to aid participant uptake. The pooled effect size on depression compared with a nonintervention control was small (Cohen *d*=0.33; 95% CI 0.11-0.55), whereas the pooled effect size of studies comparing an intervention group with an active control showed no significant differences (Cohen *d*=0.14; 95% CI −0.04 to 0.31). In addition, pooled effect sizes were higher when supervision was involved (for studies comparing digital mental health interventions with high human interaction vs no intervention: Cohen *d*=0.52, 95% CI 0.23-0.80; for studies comparing digital mental health interventions with high human interaction vs active controls with no supervision: Cohen *d*=0.49; 95% CI −0.11 to 1.01).

Web-Based Training Intervention in Behavioral Interventions and Video Training Materials

Parsons et al [35] reported that using video training materials compared with face-to-face training improved parent knowledge, parent intervention fidelity, social behavior, and communication skills of children with autism spectrum disorders.

Research Question 3: Feedback Valued by Participants With Younger Children

Feedback was valued by perinatal and postpartum women and parents in neonatal intensive care units, as reported by 4 reviews [24,28,33,34].

In the review by Merten et al [34], participants valued visual and personalized feedback, information, and tools for physical activity and dietary intake tailored to their self-monitored data during pregnancy. This feedback reinforced successes and/or offered motivational support and recommendations to achieve their goals. Ames et al [24] reported that the opportunity to offer feedback about needs, preferences, and experiences during pregnancy helped develop or improve the study intervention. In the review by Dol et al [28], parents valued a video-feedback intervention that guided them to reflect on their own successful interactions through recordings of parent-infant interaction and feedback from a video interaction guidance professional. According to Lim et al [33], many of the characteristics of DIs that postpartum women valued included feedback and goal setting. Women valued setting realistic goals through video consultation with their dietitian and tracking daily weight, exercise, and blood glucose levels in a web-based intervention, consistent with known key strategies for behavior change.

Research Question 4: Effectiveness of Engagement, Participation, and Retention Promotion Strategies

Overall, 63% (12/19) of reviews reported the effectiveness of e-engagement or participation promotion strategies. Table 2 summarizes the effectiveness of the various e-engagement, participation, and retention strategies reported in the reviews. Most findings were reported as relative rather than absolute differences, where numerical syntheses were provided (refer to Multimedia Appendix 2 [22,23,25,27,29,31,32,34,36,37,39,40] for further details on the answered research questions and strategies).

XSL•FO
RenderX

Table 2. Interventions, main outcomes, and results of included systematic reviews.

Study	Condition or sample	Intervention vs control	Outcome	Study statistics: number of studies, effect size (95% CI), heterogeneity	Results
Alkhaldi, 2016 [23]	Digital interventions	Study design features: (1) Technology-based engagement strategies (email, phone call, and SMS text messages) to promote engagement with digital interventions vs no strategy (11 studies); postal mail strategy (1 study); fewer technology-based strategies than the intervention group (2 studies).	Engagement—engagement with the digital intervention. Dichotomous outcomes: number of log-ins or visits, page views, sessions completed, digital interventions features used. Continuous outcomes: time spent on the digital intervention.	(1) • Dichotomous outcomes (n=8): RR[a]=1.27 (−1.01 to 1.60) favoring strategy group; I^2=71%.[b] • Continuous outcomes (n=4): SMD[c] 0.19 (−0.11 to 0.48) favoring strategy group; I^2=20%.	(1) Engagement in a digital intervention was higher with engagement strategy, compared with no strategy.
Baumeister, 2014 [26]	Mental health disorders	Study design features: (1) Guided interventions (with human support) vs nonguided interventions (self-guided) (2) Guided interventions with a higher qualified e-coach vs guided interventions with a lesser qualified e-coach (3) Intensive guidance (at least three email conversations per week) vs less-intensive guidance (one email contact per week).	Completion—number of completed modules, number of people completing the intervention.	(1) • Number of completed modules (n=7): SMD 0.53 (0.37 to 0.67), higher in a guided group. • Number of completers (n=6): OR[d] 2.76 (1.68 to 4.53). Higher for a guided group; I^2=42%. (2) • Number of completed modules (n=4): SMD −0.15 (−0.36 to 0.05). No significant difference between groups; I^2=0%. • Number of completers (n=4): OR 0.85 (0.54 to 1.35). No significant difference between groups; I^2=0%. (3) • Completer rate: OR 1.40 (0.41 to 4.71). Higher in intensive guidance group. • Mean completed modules: SMD 0.11 (−0.41 to 0.63). Higher in intensive guidance group.	(1) Completion was higher in guided interventions than self-guided interventions. (2) No difference in completion by coach qualification level. (3) Completion was higher with intensive guidance than less-intensive guidance.
Garrido, 2019 [30]	Internet or web-based interventions	Communication modes: (1) Web-based module learning materials or activities, group chats or courses, online forums, and web-based chat facilities with a mental health professional vs face-to-face counseling (2) Computer-based programs including games and psychoeducational computer programs vs waitlist control group.	Completion—proportion of commencing participants who completed the intervention.	(1) Not stated.	(1) and (2) Engagement and adherence rates were low with participants completing less than half of the intervention components.

XSL•FO
RenderX

Study	Condition or sample	Intervention vs control	Outcome	Study statistics: number of studies, effect size (95% CI), heterogeneity	Results
Parsons, 2017 [35]	Internet or web-based parent training programs for autism spectrum disorder	Communication modes: (1) Web-based training intervention in behavioral interventions vs written training materials (2) Video training materials vs completing the same training face-to-face within families' homes.	Completion—completion and adherence.	(1) and (2) Not stated.	(1) and (2) Interventions delivered via videos were more effective and accepted by parents than those delivered via written information.
Mertens, 2019 [34]	Telehealth	Communication modes: (1) Mobile apps, SMS text messages, and e-intervention vs standard care including brief information brochures with healthy eating and physical activity advice.	Participation—efficacy, feasibility, acceptability, use of e-intervention.	(1) Not stated.	(1) Email, app alerts, or SMS text message notifications are well accepted for health interventions.
Whitaker, 2017 [40]	Internet or web-based interventions	Communication modes: (1) Recruitment via Facebook advertisements vs recruitment via traditional methods or national data.	Engagement—number of participants recruited, conversion rate.	(1) Not stated.	(1) Facebook can be successfully used to recruit young and hard-to-reach populations. Facebook-recruited samples were generally representative to the target demographic, but some reported overrepresentation of young White women.
Atkinson, 2019 [25]	Vaccinations	Communication modes: (1) Digital push notifications (eg, SMS text message alerts) vs nondigital interventions (eg, appointment card). (2) Digital push notifications (eg, SMS text message alerts) vs nondigital pull interventions	Participation—vaccination uptake (1 dose) or completion (all doses in series).	(1) • 1 dose (n=9): OR 1.17 (1.10 to 1.23); I^2=89%. • Completion of all doses (n=4): OR 1.53 (1.13 to 2.08); I^2=82%. (2) • 1 dose or completion of all doses (n=10): OR 1.22 (1.15 to 1.30); I^2=79%.	(1) and (2) There were increased odds of participants being vaccinated or completing the vaccination series with digital alerts compared with nondigital interventions.
Dubad, 2018 [29]	Delivery mode	Communication modes: (1) Mobile mood-monitoring apps vs paper diary or in person.	Participation—completion rate of diary entries and mood assessments, engagement with the app.	(1) Not stated.	(1) Participation rates ranged between 30% and 99%.
Belisario, 2015 [27]	Delivery mode	Communication modes: (1) Smartphone app questionnaire vs paper questionnaire.	Completion—data completeness.	(1) Not stated.	(1) Higher data completeness in app than paper reported by individual studies.
Dol et al, 2017 [28]	eHealth intervention	Communication modes: (1) Videoconferencing (Skype or FaceTime), Baby CareLink (an internet-based application), video-feedback intervention, and internet-based telemedicine program vs standard care.	Completion—parents completed demographic and feasibility surveys postintervention.	(1) Not stated.	(1) Parents generally found eHealth interventions useful and acceptable for neonatal intensive unit care for their infant.

XSL•FO
RenderX

Study	Condition or sample	Intervention vs control	Outcome	Study statistics: number of studies, effect size (95% CI), heterogeneity	Results
Valimaki, 2017 [39]	Internet or web-based interventions	Communication modes: (1) Web-based interventions for depression and anxiety, computers, tablets, or mobile phones vs wait-list, other intervention method or program.	Completion—attrition, number of participants leaving the study early.	(1) • Attrition of web-based interventions compared with a control group for short-term effectiveness (n=11): OR 1.31 (1.08 to 1.58). • Attrition in midterm (follow-up measurements after 3-5 months) effectiveness (n=3): OR 1.65 (1.09 to 2.49).	(1) Adolescents in the intervention group left the study early more often, both in short-term studies and midterm studies.
Robotham, 2016 [36]	Patients attending various health care services	Contact metrics: (1) One SMS text message notification vs no SMS text message notifications. (2) 2+SMS text message notifications vs no SMS text message notification. (3) SMS text message notifications vs voice notifications.	Participation—attendance, cancellation, and "no shows" at a health care service appointment.	(1) • Attendance (n=13): RR 1.23 (1.10 to 1.38) in favor of the SMS text message group; I^2=82%. • Cancellation (n=3): RR 1.37 (*P*=.34) with no difference between groups; I^2<1%. • "No shows" (n=16): RR 0.75 (0.68 to 0.82); I^2=21%. (2) • Attendance (n=13): RR 1.49 (1.17 to 1.88) in favor of 2+notifications group; I^2=66%. 19% risk difference • "No shows": (n=15): RR 0.75 (0.57 to 0.99) with "no shows" lower in the 2+notifications group or I^2=35%. 0.3% risk difference between 1 and 2+notification groups. (3) • Attendance (n=3): RR 0.90 (0.82 to 0.98) in favor of voice notifications, I^2<1%; "No shows" (n=4): RR 1.12 (0.90 to 1.38), I^2=73% • Between 1 and 2+notification groups. "No shows": (n=15): RR 0.75 (1.17 to 1.88) with "no shows" lower in the 2+notifications group; I^2=35%. 0.3% risk difference between 1 and 2+notification groups. • Attendance (n=3): RR 0.90 (0.82 to 0.98) in favor of voice notifications, I^2<1%. "No shows" (n=4): RR 1.12 (0.90 to 1.38); I^2=73%.	(1) Patients who received SMS text message notifications were 23% more likely to attend, equally likely to cancel, and less likely to *no show* a clinic appointment than those who received no notification. (2) Participants who receive 2+SMS text message notifications are 19% more likely to attend compared with one SMS text message notification but equally likely to *not show* at a clinic appointment, compared with those who received 1 SMS text message. (3) Voice notifications may increase clinic attendance slightly compared to SMS notifications, but no difference was found for "no shows"
Thakkar, 2016 [37]	Various chronic conditions	Contact metrics: (1) SMS text message reminder vs no SMS text message reminder.	Participation—medication adherence.	(1) • Adherence to medication schedule (n=not stated): OR 2.11 (1.52 to 2.93). Weighted mean effect size (n=not stated): Cohen d=0.41 (0.23 to 0.59).	

XSL•FO
RenderX

Study	Condition or sample	Intervention vs control	Outcome	Study statistics: number of studies, effect size (95% CI), heterogeneity	Results
					(1) The odds of medication adherence more than doubled with SMS text message reminders, compared with no reminders. Assuming baseline medication adherence was 50%, this translates to an improvement to 67.8%, or an absolute increase of 17.8%.
Lattie, 2019 [32]	Participation reminders	Contact metrics: (1) Email reminders vs no email reminders.	Completion—number of sessions or assessments or prompts completed.	(1) • Number of sessions completed (n=not reported): Reminder group mean 2.9, SD 2.5; no reminder group mean 3.6, SD 2.3; t=0.88.	(1) Email reminders were not associated with completing more sessions in a web-based intervention. There were also notable rates of participant attrition and early program discontinuation in many of the studies.
Kang, 2017 [31]	Reminders	Contact metrics: (1) 7 email or SMS text message reminders vs paper appointment card.	Completion—completion of three-dose human papilloma virus vaccine schedule.	(1) • Number of people completing the vaccine schedule (n=86): email or SMS text message reminder group 34%, paper card reminder group 32%; *P*=.76.	(1) Completion rates of a vaccine schedule did not differ by reminder format (email or SMS text message, compared with paper card).
Adams, 2015 [22]	Incentives	Reimbursement and gifts or penalties: (1) Cash lottery tickets for attendance vs usual care (no incentives). (2) Loss of US $40 welfare benefits for not vaccinating vs usual care (no incentives). (3) Loss of some welfare benefits for not vaccinating vs usual care (no incentives).	Engagement or uptake—uptake of preschool vaccinations; up to date with 0-2–year vaccinations; up to date with child vaccinations.	(1) • At each follow-up time point, attendance for any reason and for vaccination was higher in incentives group. (2) • No difference in up-to-date rates at 1 or 2-years follow-up. (3) • Welfare deduction group had higher vaccination rates at 1, 2, 3, and 4 years. At age 2 years, the welfare deduction group had higher vaccine series completion.	(1) The incentives group had higher attendance. (2) No difference between the groups. (3) The welfare deduction group had higher vaccination rates.

[a]RR: relative risk.

[b]I^2 statistic: percentage of variation due to heterogeneity between studies.

[c]SMD: standardized mean difference.

[d]OR: odds ratio.

RoB for Included Systematic Reviews

The assessment of the 16 items of AMSTAR-2 from each included review is demonstrated in Multimedia Appendix 3 [22-40]; 11 systematic reviews were rated as critically low [22,23,25,26,30-36] and 8 [24,27-29,37-40] were rated as low quality. Items 7 and 10, as indicated in Table 3, were rated as particularly low quality. All systematic reviews except 1 [29] reported potential sources of conflicts of interest, including any funding they received for conducting the review, but no review reported on the sources of funding for the studies included in the review. It is important to note that AMSTAR-2 does not evaluate the quality of the primary studies. Its objective is to evaluate the methodological quality of the systematic reviews, considering how well the systematic review was conducted (eg, literature searching and data pooling). Therefore, if a systematic review included primary studies with a high RoB but the review itself was well conducted, the review tended to be rated as *high quality*.

XSL•FO
RenderX

In Table 3, we detailed the overall confidence in the results of each included systematic review. Reviews performed poorly with respect to (1) reporting sources of funding for included studies (0/19, 0%), (2) adequately investigating publication bias (small study bias) and discussing its likely impact on the results of the review (4/19, 21%), and (3) providing a list of excluded studies and justifying their exclusions (6/19, 32%).

Table 3. Overall confidence assessment (Assessing the Methodological Quality of Systematic Reviews 2 tool) of the 19 included systematic reviews.

AMSTAR-2[a] items	Yes, n (%)	Partial yes, n (%)	No, n (%)	No MA[b], n (%)
1. Did the research questions and inclusion criteria for the review include the components of PICO[c]?	13 (68)	0 (0)	6 (32)	0 (0)
2. Did the report of the review contain an explicit statement that the review methods were established before the conduct of the review and did the report justify any significant deviations from the protocol?[d]	7 (37)	8 (42)	4(21)	0 (0)
3. Did the review authors explain their selection of the study designs for inclusion in the review?	12 (63)	0 (0)	7 (37)	0 (0)
4. Did the review authors use a comprehensive literature search strategy?[d]	3 (16)	14 (74)	2 (11)	0 (0)
5. Did the review authors perform study selection in duplicates?	13 (68)	0 (0)	6 (32)	0 (0)
6. Did the review authors perform data extraction in duplicates?	12 (63)	0 (0)	7 (37)	0 (0)
7. Did the review authors provide a list of excluded studies and justify the exclusions?[d]	6 (32)	0 (0)	13 (68)	0 (0)
8. Did the review authors describe the included studies in adequate detail?	13 (68)	3 (16)	3 (16)	0 (0)
9. Did the review authors use a satisfactory technique for assessing the RoB[e] in individual studies that were included in the review?[d]	9 (47)	6 (32)	4 (21)	0 (0)
10. Did the review authors report on the sources of funding for the studies included in the review?	0 (0)	0 (0)	19 (100)	0 (0)
11. If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?[d]	7 (37)	0 (0)	4 (21)	8 (42)
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	6 (32)	0 (0)	5 (26)	8 (42)
13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?[d]	10 (53)	0 (0)	9 (47)	0 (0)
14. Did the review authors provide a satisfactory explanation for and discussion of any heterogeneity observed in the results of the review?	10 (53)	0 (0)	9 (47)	0 (0)
15. If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?[d]	4 (21)	0 (0)	8 (42)	7 (37)
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	18 (95)	0 (0)	1 (5)	0 (0)

[a]AMSTAR-2: Assessing the Methodological Quality of Systematic Reviews 2.

[b]MA: meta-analysis.

[c]PICO: population, intervention, control group, outcome.

[d]Items considered as critical domains in the AMSTAR-2.

[e]RoB: risk of bias.

Out of the 19 reviews, 10 (53%) accounted for such bias in individual studies when interpreting and discussing the results of the review and 9 (47%) reviews used a satisfactory technique for assessing the RoB in individual studies included in the review. For reviews that used a satisfactory technique for assessing RoB, most [23,26,27,29,35,36] suggested that e-engagement, participation, and retention promotion strategies were effective.

Discussion

Principal Findings

This study reviewed the current state of research comparing alternative strategies to maximize participant engagement, participation, and retention in large digital studies (as held in narrative and systematic reviews). We explored EFIs and study design features that aid engagement, participation, and retention.

XSL•FO
RenderX

For reviews that met the inclusion criteria, there was substantial heterogeneity across studies in terms of e-strategies.

Most reviews show that e-engagement and participation promotion strategies are effective, which is promising. However, these reviews canvassed relatively few experimentally tested strategies, suggesting that the myriad of alternative strategies that may have been tested have not yet been the subject of reviews. Many studies have reported these features as a secondary goal of objectives such as adherence to therapy rather than as a primary goal in and of itself. From the 19 reviews, few contained very large digital studies that directly compared alternative strategies to examine impacts on engagement and retention; this may reflect the small number of such mega-studies worldwide. However, contributing reviews contained multipurpose (observational and interventional) cohort studies conducted in clinical and research settings. Motivation for study engagement, participation, and retention may differ somewhat between observational and clinical intervention studies where the participant can potentially directly benefit from participation, although the successful strategies make sense and at face value seem likely to generalize to both settings. In the absence of more tailored evidence, engagement strategies successful in intervention studies may be the best evidence we have, though they should be cautiously applied.

In the context of technical study design features, evidence suggests that using email or SMS text message reminders and voice notifications enhanced participant attendance to health care clinics. Although promising, these results should be interpreted with caution given the short duration of the e-intervention and reliance on self-reported medication adherence measures. Future studies need to determine the features of text message interventions that improve success and appropriate patient populations, sustained effects, and influences on clinical outcomes.

Human facilitation or support was important in influencing the uptake, engagement, and outcomes of digital technologies [26]. As illustrated by completion modules and completer rates, the larger effect sizes found in guided interventions suggested increased intervention adherence.

Reviews examining the effectiveness of e-engagement, participation, and retention interventions in the context of a health care intervention (rather than a cohort study) suggested the following:

1. Visual and personalized feedback seemed effective, for example, for recordings of parent-child interaction in the neonatal intensive care setting [42,43]. This reinforces successes and/or offers motivational support to achieve an individual's goals and is consistent with known key strategies for behavior change.
2. In e-intervention studies, goal setting has mostly been used as a behavior change strategy [44-46], such as the Fishbein and Yzer Integrative Model of Behavioral Prediction and Fogg Behavior Model for Persuasive Design [44], Theory of Planned Behavior and Fun Theory [47], the Social Cognitive Theory [48,49], and the Coventry, Aberdeen, and London-Refined taxonomy of behavior change techniques [50].
3. Using digital push interventions for vaccine uptake and series completion supported the idea that digital technologies could be a useful adjunct in improving vaccination rates. Reminder interventions for vaccinations have improved the completion of vaccination schedules.
4. There was higher uptake when parental financial incentives or rewards were offered in quasi-mandatory schemes to increase the uptake of preschool vaccinations. Universal gifts were more acceptable than targeted parental financial incentives.
5. Mental health apps were effective or partially effective in producing beneficial changes in psychological outcomes among young adolescents (ie, among college students). This is consistent with past meta-analyses of digital mental health programs for similar populations [51,52].
6. Intensive guidance (with a human coach) was more efficacious than unguided interventions and a beneficial design feature, particularly for mental health studies. It is considered an adherence-facilitating measure in large digital research studies.
7. Electronic text notifications improved attendance and reduced nonattendance (no-shows) across health care settings. Sending multiple notifications improved attendance rates.

Overall, no specific e-intervention strategy was identified as being superior. However, more interactive methods of delivery, such as videos and regular e-therapist contact for training, (1) improve adherence, (2) increase completion rates, and (3) improve fidelity. Further research is needed to understand the strategies that improve retention in longitudinal studies.

Limitations

We limited our search to systematic reviews published between 2012 and 2019. These reviews should give good reach into source studies during the preceding decade while encompassing the rapid evolution of technology and the explosion of digital methods in this period and thus relevant to the new studies of the 2020s. However, we acknowledge that this is an arbitrary choice. Although all of the literature sourced reported on studies using partial or fully digital contact with participants, much was in the context of interventions and may not be wholly applicable to observational cohort studies. Nonetheless, those strategies found to be successful in interventional settings seem worthwhile to explore in cohort studies. We obtained low-quality ratings for some systematic reviews. We also note that although high engagement and retention are the best strategies to obtain powerful representative data sets, statistical techniques such as multiple imputation are vital adjuncts.

Conclusions

Although all studies want to maximize the recruitment and retention of study participants, the best methods to do this, particularly in digital settings, are understudied. This review adds to the small but growing literature on methods for optimizing engagement and participation in digital contact cohort studies. Evidence-based recruitment and retention methods are particularly important to the success of the next generation of very large birth cohorts, which are very expensive but have low funding per participant and require high retention

XSL•FO
RenderX

throughout decades despite participants having no or very little in-person contact with the study team. Ideally, such studies will not only use existing evidence-based methods but will also build on experimental studies of alternative engagement and retention methods to build the evidence base of *the science of science*.

Acknowledgments

This study was funded by GenV. GenV is supported by the Paul Ramsay Foundation, the Victorian Government, the Murdoch Children's Research Institute, and the Royal Children's Hospital Foundation (2019-1226). Research at the Murdoch Children's Research Institute was supported by the Victorian Government's operational infrastructure support program.

MW is supported by Australia's National Health and Medical Research Council (NHMRC) Principal Research Fellowship 1160906. FKM is supported by NHMRC Career Development Fellowship 1111160.

Authors' Contributions

MW conceived the study. MW, JN, and SAC designed the study. JN, SAC, YW, and LC conducted the review. JN drafted the first version of this manuscript. All authors contributed to writing the manuscript and read and approved the final review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search query for systematic reviews and meta-analyses in OVID MEDLINE (search conducted on December 19, 2019).
[PDF File (Adobe PDF File), 166 KB - jmir_v23i5e23499_app1.pdf]

Multimedia Appendix 2
Study-specific research questions and answers provided in reviews to achieve high participant engagement, participation, and retention.
[PDF File (Adobe PDF File), 544 KB - jmir_v23i5e23499_app2.pdf]

Multimedia Appendix 3
Methodological quality of included systematic reviews.
[PDF File (Adobe PDF File), 578 KB - jmir_v23i5e23499_app3.pdf]

References

1. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol 2017 Nov 01;186(9):1026-1034 [FREE Full text] [doi: 10.1093/aje/kwx246] [Medline: 28641372]
2. McCarthy M. US cancels plan to study 100,000 children from "womb" to age 21. Br Med J 2014 Dec 19;349:g7775. [doi: 10.1136/bmj.g7775] [Medline: 25526900]
3. Pearson H. Massive UK baby study cancelled. Nature 2015 Oct 29;526(7575):620-621. [doi: 10.1038/526620a] [Medline: 26511556]
4. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C, MoBa Study Group. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol 2006 Oct;35(5):1146-1150. [doi: 10.1093/ije/dyl170] [Medline: 16926217]
5. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
6. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
7. 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 Interventions (EEPI) Randomized Controlled Trial. Contemp Clin Trials 2019 Mar;78:107-115. [doi: 10.1016/j.cct.2019.01.015] [Medline: 30711664]
8. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

XSL•FO
RenderX

9. Teague S, Youssef GJ, Macdonald JA, Sciberras E, Shatte A, Fuller-Tyszkiewicz M, SEED Lifecourse Sciences Theme. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Med Res Methodol 2018 Nov 26;18(1):151 [FREE Full text] [doi: 10.1186/s12874-018-0586-7] [Medline: 30477443]
10. Gustavson K, von Soest T, Karevold E, Røysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health 2012 Oct 29;12:918 [FREE Full text] [doi: 10.1186/1471-2458-12-918] [Medline: 23107281]
11. Szklo M. Population-based cohort studies. Epidemiol Rev 1998;20(1):81-90. [doi: 10.1093/oxfordjournals.epirev.a017974] [Medline: 9762511]
12. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health 2011 Apr 19;11:249 [FREE Full text] [doi: 10.1186/1471-2458-11-249] [Medline: 21504610]
13. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008 Jun;93(6):458-461. [doi: 10.1136/adc.2007.127316] [Medline: 18495909]
14. Siddiqi A, Sikorskii A, Given CW, Given B. Early participant attrition from clinical trials: role of trial design and logistics. Clin Trials 2008;5(4):328-335 [FREE Full text] [doi: 10.1177/1740774508094406] [Medline: 18697847]
15. Gupta A, Calfas KJ, Marshall SJ, Robinson TN, Rock CL, Huang JS, et al. Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a Marketing and Information Technology (MARKIT) model. Contemp Clin Trials 2015 May;42:185-195 [FREE Full text] [doi: 10.1016/j.cct.2015.04.002] [Medline: 25866383]
16. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. BMC Med Res Methodol 2006 Jul 19;6:34 [FREE Full text] [doi: 10.1186/1471-2288-6-34] [Medline: 16854229]
17. Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol 2007 Aug;60(8):757-765 [FREE Full text] [doi: 10.1016/j.jclinepi.2006.11.023] [Medline: 17606170]
18. Robinson KA, Dinglas VD, Sukrithan V, Yalamanchilli R, Mendez-Tellez PA, Dennison-Himmelfarb C, et al. Updated systematic review identifies substantial number of retention strategies: using more strategies retains more study participants. J Clin Epidemiol 2015 Dec;68(12):1481-1487 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.013] [Medline: 26186981]
19. Wake M, Hu YJ, Warren H, Danchin M, Fahey M, Orsini F, et al. Integrating trials into a whole-population cohort of children and parents: statement of intent (trials) for the Generation Victoria (GenV) cohort. BMC Med Res Methodol 2020 Sep 24;20(1):238 [FREE Full text] [doi: 10.1186/s12874-020-01111-x] [Medline: 32972373]
20. 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]
21. 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]
22. Adams J, Bateman B, Becker F, Cresswell T, Flynn D, McNaughton R, et al. Effectiveness and acceptability of parental financial incentives and quasi-mandatory schemes for increasing uptake of vaccinations in preschool children: systematic review, qualitative study and discrete choice experiment. Health Technol Assess 2015 Nov;19(94):1-176 [FREE Full text] [doi: 10.3310/hta19940] [Medline: 26562004]
23. 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 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
24. Ames HM, Glenton C, Lewin S, Tamrat T, Akama E, Leon N. Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis. Cochrane Database Syst Rev 2019 Oct 14;10:CD013447 [FREE Full text] [doi: 10.1002/14651858.CD013447] [Medline: 31608981]
25. Atkinson KM, Wilson K, Murphy MS, El-Halabi S, Kahale LA, Laflamme LL, et al. Effectiveness of digital technologies at improving vaccine uptake and series completion - a systematic review and meta-analysis of randomized controlled trials. Vaccine 2019 May 21;37(23):3050-3060. [doi: 10.1016/j.vaccine.2019.03.063] [Medline: 31036457]
26. 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]
27. Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, 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. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
28. Dol J, Delahunty-Pike A, Anwar Siani S, Campbell-Yeo M. eHealth interventions for parents in neonatal intensive care units: a systematic review. JBI Database System Rev Implement Rep 2017 Dec;15(12):2981-3005. [doi: 10.11124/JBISRIR-2017-003439] [Medline: 29219875]

XSL•FO
RenderX

29. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]
30. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
31. Kang HS, De Gagne JC, Son YD, Chae S. Completeness of human papilloma virus vaccination: a systematic review. J Pediatr Nurs 2018;39:7-14. [doi: 10.1016/j.pedn.2017.12.003] [Medline: 29525217]
32. 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]
33. Lim NL, Shorey S. Effectiveness of technology-based educational interventions on the empowerment related outcomes of children and young adults with cancer: a quantitative systematic review. J Adv Nurs 2019 Oct;75(10):2072-2084. [doi: 10.1111/jan.13974] [Medline: 30740758]
34. Mertens L, Braeken MA, Bogaerts A. Effect of lifestyle coaching including telemonitoring and telecoaching on gestational weight gain and postnatal weight loss: a systematic review. Telemed J E Health 2019 Oct;25(10):889-901. [doi: 10.1089/tmj.2018.0139] [Medline: 30523742]
35. Parsons D, Cordier R, Vaz S, Lee HC. Parent-mediated intervention training delivered remotely for children with autism spectrum disorder living outside of urban areas: systematic review. J Med Internet Res 2017 Aug 14;19(8):e198 [FREE Full text] [doi: 10.2196/jmir.6651] [Medline: 28807892]
36. Robotham D, Satkunanathan S, Reynolds J, Stahl D, Wykes T. Using digital notifications to improve attendance in clinic: systematic review and meta-analysis. BMJ Open 2016 Oct 24;6(10):e012116 [FREE Full text] [doi: 10.1136/bmjopen-2016-012116] [Medline: 27798006]
37. 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;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
38. Tromp K, Zwaan CM, van de Vathorst S. Motivations of children and their parents to participate in drug research: a systematic review. Eur J Pediatr 2016 May;175(5):599-612 [FREE Full text] [doi: 10.1007/s00431-016-2715-9] [Medline: 27041121]
39. Välimäki M, Anttila K, Anttila M, Lahti M. Web-based interventions supporting adolescents and young people with depressive symptoms: systematic review and meta-analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]
40. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
41. Safran C, Pompilio-Weitzner G, Emery KD, Hampers L. Collaborative approaches to e-Health: valuable for users and non-users. Stud Health Technol Inform 2005;116:879-884. [Medline: 16160369]
42. Brown MJ, Sinclair M, Liddle D, Hill AJ, Madden E, Stockdale J. A systematic review investigating healthy lifestyle interventions incorporating goal setting strategies for preventing excess gestational weight gain. PLoS One 2012;7(7):e39503 [FREE Full text] [doi: 10.1371/journal.pone.0039503] [Medline: 22792178]
43. Gardner B, Wardle J, Poston L, Croker H. Changing diet and physical activity to reduce gestational weight gain: a meta-analysis. Obes Rev 2011 Jul;12(7):602-620. [doi: 10.1111/j.1467-789X.2011.00884.x] [Medline: 21521451]
44. Fernandez ID, Groth SW, Reschke JE, Graham ML, Strawderman M, Olson CM. eMoms: electronically-mediated weight interventions for pregnant and postpartum women. Study design and baseline characteristics. Contemp Clin Trials 2015 Jul;43:63-74 [FREE Full text] [doi: 10.1016/j.cct.2015.04.013] [Medline: 25957183]
45. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
46. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized mobile health intervention for health and weight loss in postpartum women receiving women, infants, and children benefit: a randomized controlled pilot study. J Womens Health (Larchmt) 2017 Jul;26(7):719-727 [FREE Full text] [doi: 10.1089/jwh.2016.5947] [Medline: 28338403]
47. Kernot J, Olds T, Lewis LK, Maher C. Effectiveness of a facebook-delivered physical activity intervention for post-partum women: a randomized controlled trial protocol. BMC Public Health 2013 May 29;13:518 [FREE Full text] [doi: 10.1186/1471-2458-13-518] [Medline: 23714411]
48. Phelan S, Brannen A, Erickson K, Diamond M, Schaffner A, Muñoz-Christian K, et al. 'Fit Moms/Mamás Activas' internet-based weight control program with group support to reduce postpartum weight retention in low-income women: study protocol for a randomized controlled trial. Trials 2015 Mar 25;16:59 [FREE Full text] [doi: 10.1186/s13063-015-0573-9] [Medline: 25887964]

XSL•FO
RenderX

49. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014 Nov;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
50. Willcox J, Wilkinson S, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. Brit J Obstet Gynec 2017 Oct 20;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
51. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
52. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]

Abbreviations

AMSTAR-2: Assessing the Methodological Quality of Systematic Reviews 2
DI: digital intervention
EFI: engagement-facilitation intervention
GenV: Generation Victoria
NHMRC: National Health and Medical Research Council
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB: risk of bias

Edited by G Eysenbach; submitted 14.08.20; peer-reviewed by S Teague, D McDonnell; comments to author 23.10.20; revised version received 16.03.21; accepted 07.04.21; published 14.05.21.

Please cite as:
Nkyekyer J, Clifford SA, Mensah FK, Wang Y, Chiu L, Wake M
Maximizing Participant Engagement, Participation, and Retention in Cohort Studies Using Digital Methods: Rapid Review to Inform the Next Generation of Very Large Birth Cohorts
J Med Internet Res 2021;23(5):e23499
URL: https://www.jmir.org/2021/5/e23499
doi:10.2196/23499
PMID:33988509

©Joanna Nkyekyer, Susan A Clifford, Fiona K Mensah, Yichao Wang, Lauren Chiu, Melissa Wake. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.05.2021. 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

Use of Twitter in Neurology: Boon or Bane?

Biswamohan Mishra[1], MD; Monica Saini[2], MD; Carolynne M Doherty[3], MBChB, MRCP; Robert D S Pitceathly[3], MBChB, PhD, MRCP; Roopa Rajan[1], MD, DM; Omar K Siddiqi[4,5], MD, MPH; Gita Ramdharry[3], PhD; Ajay Asranna[6], MD, DM; Pedro Jose Tomaselli[7], MD, MSc, PhD; Allan G Kermode[8], MBBS, MD, FRACP, FRCP; Jawad A Bajwa[9], MD; Divyani Garg[10], MD, DM; Venugopalan Y Vishnu[1], MD, DM

[1]Department of Neurology, All India Institute of Medical Sciences, New Delhi, India

[2]Department of Neurology, National Neuroscience Institute, Novena, Singapore

[3]Department of Neuromuscular Diseases, University College London Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, University College London, London, United Kingdom

[4]Department of Neurology, Harvard Medical School, Boston, MA, United States

[5]Department of Neurology, University of Zambia School of Medicine, Lusaka, Zambia

[6]Department of Neurology, National Institute of Mental Health and Neurosciences, Bengaluru, India

[7]Department of Neurosciences and Behaviour Sciences, Clinical Hospital of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil

[8]Department of Neurology, Perron Institute, Queen Elizabeth II Medical Centre, University of Western Australia, Perth, Australia

[9]Department of Neurology, National Neuroscience Institute, King Fahad Medical City, Riyadh, Saudi Arabia

[10]Department of Neurology, Lady Hardinge Medical College, New Delhi, India

Corresponding Author:
Venugopalan Y Vishnu, MD, DM
Department of Neurology
All India Institute of Medical Sciences
Room No. 704, CN Centre
New Delhi, 110029
India
Phone: 91 9855480361
Email: vishnuvy16@yahoo.com

Abstract

Twitter is a free, open access social media platform that is widely used in medicine by physicians, scientists, and patients. It provides an opportunity for advocacy, education, and collaboration. However, it is likely not utilized to its full advantage by many disciplines in medicine, and pitfalls exist in its use. In particular, there has not been a review of Twitter use and its applications in the field of neurology. This review seeks to provide an understanding of the current use of Twitter in the field of neurology to assist neurologists in engaging with this potentially powerful application to support their work.

(J Med Internet Res 2021;23(5):e25229) doi:10.2196/25229

KEYWORDS

Twitter; neurology; tweet chats; research; tweetorials; contemporary issues

Introduction

Twitter is a free-to-use social networking and microblogging platform, enabling registered users to post (ie, tweet), read, and repost (ie, retweet) short messages known as tweets. Tweets are limited to 280 characters—previously 140—and users can share photos, short videos, or website URLs [1]. The limited character count is an important consideration when tweeting, and encourages succinct delivery of concepts or acts as the starting point for a conversation. Messages from profiles the user has *followed*, using their @handles, appear in a curated *feed*, which is updated with new tweets at different intervals. Though Twitter has been widely used in medicine by physicians, scientists, and patients, its application in the field of neurology remains relatively uncharted.

Here we explore the powerful potential of Twitter in promoting research, education, and health care activities in neurology to relevant stakeholders. We also examine the risks and limitations of Twitter use in a professional capacity.

XSL•FO
RenderX

Search Strategy and Selection Criteria

References for this review were identified by searching PubMed, Google Scholar, Embase, MEDLINE, and the Twitter site between March 2006 (ie, when Twitter was founded) and July 2020 as well as searching references from relevant articles. The search terms "Twitter," "neurology," "journal clubs," "tweetorials," "tweet chats," "misuse," "unprofessional," "social media," "health care," and "medicine" were used. The search was restricted to English-language articles. The final reference list was generated on the basis of relevance to the topics covered in this review.

Neurology and Twitter

Twitter has become an appealing platform for the scientific community to exchange ideas, information, research insights, teaching and career achievements, and opportunities. Many academic Twitter hot spots garner a huge following (eg, #AcademicChatter, #PhDChat, #Medtwitter, @OpenAcademics, and #Epitwitter). As with other social media platforms, neurologists, neurology residents, neuroscience researchers, advocacy groups, journals, professional societies, and funding agencies have all asserted their presence in the *Twitterverse* over the past few years (Tables 1 and 2 [2-4]). Tweets addressed to neurologists are often tagged #neurotwitter, and users might post anything ranging from introductions to recently published papers from their institutions to pictures of electroencephalogram (EEG)–adorned cakes.

Table 1. Selected neurology journals, professional associations, and medical funding agencies with a strong Twitter presence.

Journal, professional association, or funding agency	Twitter handle	Approximate follower count
Journal		
Green Journal	@GreenJournal	44,900
Lancet Neurology	@TheLancetNeuro	26,900
American Academy of Neurology (ANN)	@AANMember	23,200
Brain	@Brain1878	19,800
Nature Reviews Neurology	@NatRevNeurol	28,100
Nature Neuroscience	@NatureNeuro	86,400
JAMA Neurology	@JAMANeuro	35,700
Neurology Today	@NeurologyToday	44,000
Association		
Motor Neuron Disease (MND) Association	@mndassoc	31,000
Stroke Association	@TheStrokeAssoc	107,000
Alzheimer's Research UK	@AlzResearchUK	78,000
American Stroke Association	@American_Stroke	19,100
Brain and Life Magazine	@BrainandLifeMag	34,300
Dementia UK	@DementiaUK	106,200
International Parkinson and Movement Disorder Society	@movedisorder	7157
World Muscle Society	@WorldMuscleSoc	800
The Peripheral Nerve Society (PNS)	@PNSociety1	200
Funding agency		
National Institutes of Health (NIH)	@NIH	1,100,000
National Institute of Neurological Disorders and Stroke (NINDS)	@NINDSnews	9555
Medical Research Council (MRC)	@The_MRC	65,400
Indian Council of Medical Research (ICMR)	@ICMRDELHI	126,400

XSL•FO
RenderX

Table 2. Examples of patient support groups for neurological disorders that have a Twitter presence.

Support group	Support group Twitter handle	Activity
The Brain Charity	@TheBrainCharity	Offers emotional support, practical help, and social activities to anyone with a neurological condition
Multiple Sclerosis Foundation	@MS_Focus [3]	Helps people with multiple sclerosis (MS) live their best lives through free programs, services, and education
Stroke Association	@TheStrokeAssoc	Supports people in rebuilding their lives after stroke with their #RebuildingLives campaign
Alzheimer's Society	@alzheimerssoc	Provides information and support for people affected by dementia (#UnitedAgainstDementia)
FSHD Society	@FSHDSociety [4]	A research-focused patient organization for facioscapulohumeral muscular dystrophy (FSHD)
Amyloidosis Support Groups	@AmyloidosisSupp	Official Twitter account of Amyloidosis Support Groups; dedicated to the support of amyloidosis patients, caregivers, and former caregivers
Parkinson's Foundation	@ParkinsonDotOrg	Strives to make life better for people with Parkinson disease through expert care and research
FND Action	@FNDAction	A charity organization for raising awareness of functional neurological disorders (FNDs) [2], including nonepileptic attack disorder (#NEAD), and supporting those diagnosed with FNDs (#action4FND)
Muscular Dystrophy Association	@MDAorg	Strives to transform the lives of people affected by muscular dystrophy (MD), amyotrophic lateral sclerosis, and related neuromuscular diseases through innovations in science and in care
Friedreich's Ataxia Research Alliance	@CureFA_org	Works to direct and focus the resources and relationships needed to treat and cure Friedreich's Ataxia (FA)
Hereditary Neuropathy Foundation	@CMTNeuropathy	Supports patients and families with updated relevant information and is dedicated to funding Charcot-Marie-Tooth disease (CMT)–related research

Promoting Research

Enhance Research Visibility and Recruitment

Twitter hashtags (ie, #) help connect people with topics of interest. By following specific profiles and hashtags and reading the associated threads, one can quickly establish an overview of the most up-to-date research activities, clinical advances, and future perspectives. This effect is particularly apparent in rare diseases, where following key handles (ie, @) and hashtags ensures that researchers, physicians, and patients can establish connections; for example, #ALSResearch (amyotrophic lateral sclerosis, ALS), @Phelan_McDermid and #PMSF (Phelan McDermid Syndrome Foundation), @thetinman_org (help and awareness for stiff-person syndrome), @yaya4HL (Yaya Foundation for 4H leukodystrophy, 4HL), @fragilexsyndrom, and numerous others. Such social media cross talk is often helpful in building clinical registries and biorepositories and may facilitate higher-quality clinical and translational research.

Twitter can help promote the visibility of clinical trials, enhancing both recruitment and inclusivity via links to enrollment websites [5]. Some such examples are as follows:

1. The Emergency Laparotomy and Frailty (ELF) study achieved a target enrollment of 500 participants by utilizing a Twitter handle (@ELFStudy) and eye-catching logos on its Twitter profile [6].
2. The AVERT DOSE trial (@AVERTDOSEtrial) posts regular updates on their work evaluating rehabilitation after stroke via their Twitter profile [7], ensuring sustained visibility.
3. Novartis is recruiting participants with advanced malignancies for US clinical trials, and its Twitter profile (@NovartisOncCT) shares the following link for potential participants to join: "http://bit.ly/2LhfJIk! #ClinicalTrialsSM" [8].
4. Wasilewski et al, in the context of a caregiver study, conducted a secondary analysis of their Twitter recruitment and found that out of the 71 caregivers, 27 were recruited via Twitter. General recruitment tweets were most frequently shared by users [9].

In a study by Sedrak et al, while the majority of the tweets (86.3%) had embedded links to news articles, one tweet led to a patient recruitment site (ie, ClinicalTrials.gov; row 7 of the second table of the study) [5]. Harnessing Twitter in this way can result in novel and larger networks.

Facilitate International Connections, Networks, and Visibility

Twitter enables users from around the world to interact with individuals at the forefront of academic research or clinical practice. This interaction may be via tweets, comments, or both—an engagement that might not otherwise have been possible. Conference organizers can also enhance the influence of conference-related discussions beyond those attending in person through social media, including Twitter [10]. Following the people you met at a conference can also be a convenient way of light-touch networking long after a conference has ended.

XSL•FO
RenderX

Some fellowship training programs use Twitter to advertise; in the United States, neurosurgery programs have the highest social media presence compared to neurology and neuroradiology [11], especially in terms of advertising residency programs [12]. Since the onset of the COVID-19 pandemic, there has been at least a 216% (from n=24 to n=76) increase in the number of neurology residency accounts on Twitter, as noted by Zelikovich et al in a blog post in Neurology [13].

Within specialties, there are focused spaces for discussion on specific topics or diseases—a global connection for users with a small area of shared interest. For instance, @Microbleeds discusses microbleed research and @ScienceofPD highlights Parkinson disease research [14]. Many also use Twitter to disseminate collaborative subspecialty education, such as neuropathology and neuroradiology, to broader audiences.

Rapid Communication of Research

Twitter provides a platform where opinions, queries, and comments on recently published research can be posted, and reacted to, in real time facilitating easy access for clinical neurologists to recent advances. Examples of highly followed journals and societies can be seen in Table 1. Connecting clinicians and researchers is vital, as research data with translational potential may be helpful in management decisions. Moreover, input from physicians may prompt research in novel and impactful fields. Patient or public comments on research topics may also provide new perspectives and relevant research questions. Thus, Twitter can facilitate mutually beneficial, multichannel communication [15].

Capturing a Wider Audience, Faster: #Power

Twitter facilitates patient involvement in care and has the capacity to remove barriers to understanding complex medical conditions through sharing information. By using Twitter, a patient can connect to support groups pertaining to his or her illness, such as the Motor Neuron Disease (MND) Association (@mndassoc) [16] and Parkinson's Foundation (@ParkinsonDotOrg) [17]. Patients may interact with others who have the same disease, share experiences, and discuss treatment facilities and trials, advances in disease management, prevention, and potential sources of financial support, all of which enhance patient care. Twitter also provides a forum for all parties vested in health care to interact, including physicians, allied health care providers, patients, carers, advocacy groups, policy makers, and the pharmaceutical industry. Tagging using hashtags (ie, #) helps identify and, therefore, rapidly disseminate information to others, creating a mechanism for online multidisciplinary and interdisciplinary feedback, which may improve coordination among all relevant health care stakeholders [18].

Role as a Search Engine: Hashtags and Handles

Twitter allows for focused reading on a particular topic of interest by providing links to research articles. For instance, someone interested in clinical genetics can follow #GeneTherapy, browse through tweets, and obtain links to pertinent websites, such as the following tweet:

> *Standardizing Cell and #GeneTherapy Remains Bioindustry Goal -- Please Read: http://ow.ly/3fJu50AQgLP; @AVROBIO about characterizing stem cells for gene therapy.*

Individuals who have or research a specific disease can keep up to date at the level of the individual gene; for example, the inherited neuropathy gene PMP22 has its own hashtag: #PMP22.

Conference Updates and Participation

Conference updates are shared through individual Twitter handles (eg, @SfN2020 and @abnconf) and via broader platforms using hashtags (eg, #BBIConf2018, #TheNeuroNetwork, #NeuroConference, and #AANAM). For future conferences, links can be utilized for abstract submission (eg, #Neurology2021), resulting in broader coverage and reach. It is possible to share the latest research presented during and after a conference, increasing accessibility to those unable to travel and attend in person. Consider a neurologist at a conference who replies to a tweet from @GreenJournal; others may respond in real time with critiques and may share related papers, real-life experiences, and links to related conference presentations or other research. This intertwining of experience, expertise, and communication has the potential to advance the reach of health care delivery systems. Virtual conferences have found renewed popularity in the context of the COVID-19 pandemic and benefit from advertising via Twitter; for instance, @WorldStrokOrg recently tweeted the following:

> *#ESOWSO2020 will run as a fully virtual event from 7-9 November. Virtual sessions will allow live interaction and will be available on-demand in the weeks following the Conference.*

Another example from the Movement Disorder Society (MDS) Twitter handle @movedisorder is as follows:

> *New in #MDSMovingAlong: Looking Forward to the First MDS Virtual Congress. Read more from the MDS Virtual Congress Task Force on plans for this year's big event. There is still time to register at no charge before it begins on September 11. #MDSCongress.*

Platform for Journal Club and Research Review

Journal clubs on Twitter are an emerging alternative to in-person journal clubs, which are limited by time or geographical location. Twitter provides accessibility and flexibility to involve people globally, without the added pressure of hierarchy, ensuring a broader audience. For example, #NCSTJC is the Twitter hashtag for the Neurocritical Care Society's journal club. Similar to an in-person journal club, at a Twitter journal club, a paper is selected, sometimes using crowdsourcing to ascertain participant interest. At a specific time, participants log on to Twitter, and there may be a facilitator similar to an in-person setting. Preselected questions and polls may be posted via handles or hashtags to the participants. Though there is potential for textual discussion, it is limited by the 280-character limit, unless a thread is used [19].

Traditionally, journals provide a restricted and moderated platform for research review and comments on recently

XSL•FO
RenderX

published data. Twitter provides a more open platform for instant, multidimensional feedback on research quality and validity. On an individual level, reading through such comments may help broaden knowledge and enhance analytical skills. On a larger canvas, it may identify research flaws and knowledge gaps, ultimately leading to better designed and focused research. One can even seek comments from a particular organization or individual using their Twitter handle. Furthermore, there is a distinct advantage of obtaining varied views across the globe for one's research, rather than the tunnel vision that can exist within smaller communities.

An Alternative to Publishing Powerhouses

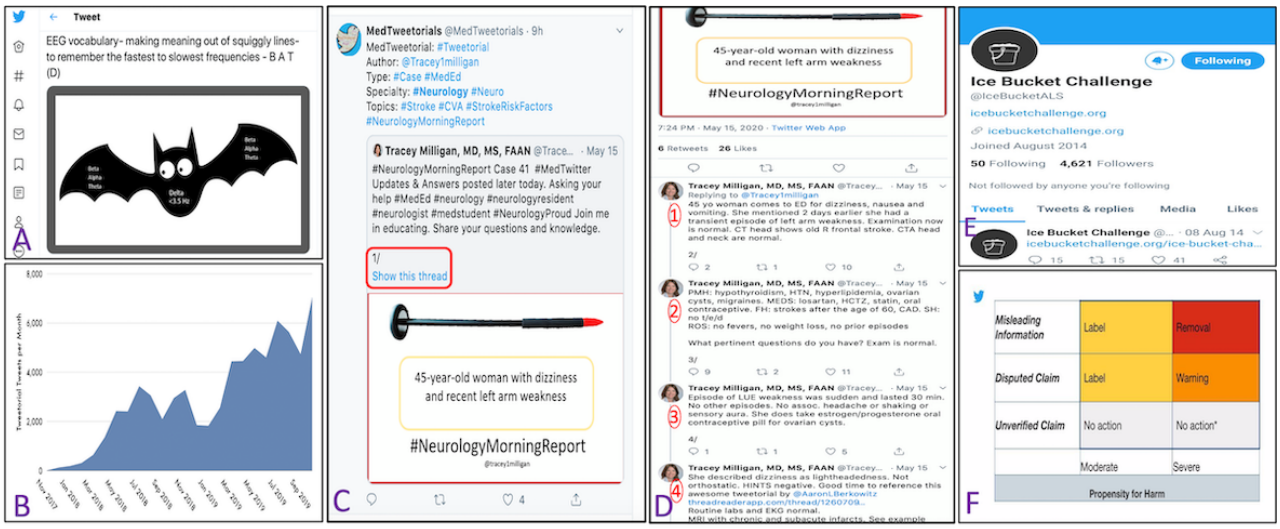
Academics and the public can be biased toward publications in high-impact journals, which may result in excellent research work that is published in less "prestigious" journals, thereby gaining less visibility. Social media outlets, such as Twitter, offer individuals and organizations such as universities a platform to publicize their work. However, expression of one's expertise and research work through social networking sites, instead of through conventional platforms like journals, can be considered as less thorough or pertinent [20].

Education and Learning

Case Discussions

Twitter has opened up new dimensions to education and learning. For example, Tracy Milligan (@Tracey1milligan), through #NeurologyMorningReport [21], shares clinically relevant neurology cases via interactive, stepwise reasoning. Twitter provides a platform to tweet bite-sized informative pearls, and by the virtue of the opportunity for interaction, the chance to learn from each other's varied clinical expertise and experience. Short, snappy, and humorous posts (eg, picture of a bat as a tool to remember EEG wave frequencies; Figure 1, A) result in a fun-filled learning experience. Thus, Twitter enhances traditional teaching by making it more interactive, exciting, and practical.

Figure 1. Examples of various aspects of Twitter that can be applied to neurologists. A. A representative tweet depicting a pictorial mnemonic on how to remember electroencephalogram (EEG) wave frequencies shared on Twitter. B. Monthly volume of tweetorials showing exponential growth; data were taken from Healthcare Social Graph using keyword and hashtag versions of tweetorials from November 2017 to October 2019. C and D. Examples of the use of tweetorials in neurology. C. Tracey Milligan shares an interesting clinical scenario and clicking on "show this thread" (red rectangle) opened up all the discussions pertinent to this case (D, 1), numbered here as 1, 2, 3, 4, and so on. E. A snapshot of the Ice Bucket Challenge movement on Twitter that was started to garner donations for amyotrophic lateral sclerosis (ALS) research. F. Brightly colored labels that Twitter is contemplating adding directly beneath lies and misinformation posted by users.



Tweetorials

Developed in late 2017 by Mike Thompson and Vinay Prasad, *tweetorials* have seen explosive growth (Figure 1, B) [22] and have recently been accepted by the New England Journal of Medicine as a type of publication under the handle @MedTweetorials [23]. A tweetorial is a series of educational tweets posted in quick succession or a long or extended tweet, made by stringing together a series (ie, thread) of tweets [24]. Stringing together tweets, or composing a thread, allows the educator to get around the 280-character limit per tweet [25]. Because they are typically *tagged* with a keyword or hashtag and the tweets numbered by the convention 1/n, tweetorials are easy to retrieve (Figure 1, C and D). As clinical reasoning forms the backbone of neurology teaching, the structure of tweetorials has become immensely popular. For example, Crystal Yeo (@CrystalYeoMDPhD) shares tutorials on neuromuscular conditions as tweetorials.

Tweet Chats

Tweet chats are public Twitter conversations around a unique hashtag. This hashtag allows one to track and participate in discussions as well as to connect with health care personnel and laypersons with specific interests [26]. For instance, the World Stroke Organization's (@WorldStrokeOrg) campaign through tweet chats on COVID-19 helped spread much-needed awareness on challenges and workflow modifications during the pandemic. A representative tweet is as follows:

> *In the meantime, mark your calendar for the next #WSOwebinar on August 18: #Stroke Guidelines in the setting of #COVID19 with @patrice_lindsay,*

XSL•FO
RenderX

@sheilambrasiland Jeyaraj Pandian, @gsaposnik @neurologija @WorldStrokeEd @IntJStroke @WStrokeCampaign @BelsonSarah.

Recently, NeurologyLive and the Women Neurologists Group hosted a tweet chat on mental health in neurology, with topics addressing mental health issues and available resources from the clinician's perspective; relevant tweets were tagged with #MindMoments. Recent tweets concerning #MindMoments have generated a staggering 1.1 million impressions [27].

Learning and Training Opportunities

Twitter can be used to advertise and access training opportunities across all aspects of neuroscience. Examples include @NeurologistJobs, @ABNTrainees, @GetNeurologJobs, and, for the US match system, @NMatch2021. Scholars can identify and connect to potential mentors. Users of the #PhDChat network [28], created by UK doctoral students (>12,000 followers), offer advice on identifying best-fit training programs, exchange resources, and extend support via sharing experiences in academia. During the COVID-19 pandemic, PhD writing rooms, such as those hosted by @PhDForum, have evolved to help candidates focus. Curating one's feed by following those providing relevant material can result in daily educational snippets with minimal effort. Some notable examples are as follows: (1) case- and image-based discussion on EEG hosted by Rebecca Fasano, @RebeccaFasanoMD (eg, Emory EEG Case of the Day #79 [29]); (2) @TheNotoriousEEG hosted by M Scott Perry MD [30]; (3) @ClinicalNeuroph [31], which provides news from Clinical Neurophysiology, Clinical Neurophysiology Practice, and International Federation of Clinical Neurophysiology; and (4) @MDedgeNeurology [32], which gives an overview of the most recent research and clinical developments in neurology and neuroscience.

Workplace and Professional Concerns

Work-life balance and healthy working environments are prioritized by health care institutions. Recent studies have shown burnout to be very common among all neurology practice settings and subspecialties, with one study reporting that 60% of respondents had at least one symptom of burnout [33], and another study reporting burnout to be as high as 73% and 55% among neurology residents and fellows, respectively [34]. A study on Canadian physicians reported that more than a quarter of the physicians had mental health concerns, mostly depressed mood [35]. However, sharing of professional or workplace challenges is not always straightforward and should be approached with careful judgement and with the social media policies of relevant organizations in mind. Twitter provides a platform for individuals to connect with others with shared experiences or concerns and enables discussion that is centered on coping strategies and suggestions; the platform can act as an online *professional support group* of sorts. Examples include #BlackInNeuro, @WNGtweets (Women Neurologists Group), @LGBT_PsychNeuro, and others. Such groups have the potential to reduce isolation and enhance connectivity and visibility, and act as one of the spaces for people to discuss how to best manage the inherent challenges of the working and academic worlds.

Community Health Care–Related Activities

Support Groups

Many community-centric health care activities are shared through the Twitterverse. Support groups for a variety of common and, importantly, rare neurological disorders enable relationships beyond the traditional physical space (Table 2).

Awareness Campaigns

Public awareness campaigns that use Twitter can reach a broader audience within a shorter time frame than more traditional advertising platforms. For instance, the ALS Ice Bucket Challenge (@IceBucketALS) [36] was conceived to promote awareness of ALS and to encourage donations to research. It went on to become *viral* [37] on social media during July and August 2014 (Figure 1, E), resulting in donations of US $227.6 million for ALS research in 2014 and the subsequent 4 years [38].

A study conducted in 2011 uncovered the stigma associated with epilepsy on Twitter, as 41% of tweets related to "seizure" were seen to be derogatory in nature [39]. Similar studies have also noted the use of pejorative tweets referring to psychogenic nonepileptic seizures, even by professionals [40]. Sirven et al called for a Twitter revolution to alter the way in which we perceive seizures [41]. Today, many epilepsy patient advocacy groups like the Epilepsy Foundation of America (@EpilepsyFdn; >34,000 followers) and Epilepsy Action (@epilepsyaction; >28,000 followers) regularly tweet content aiming to raise public awareness and knowledge of seizures. Only a few years later in their 2017 analysis, Meng et al found that within epilepsy and seizure accounts, verified foundations and support groups had the greatest number of accounts and users on Twitter and Facebook, with 48% of their posts aiming to provide information or to dispel common misconceptions and stigma surrounding epilepsy [42].

Patients as a Part of Focus Groups

Twitter has been adopted by patient-focused groups (eg, NeuroImmunology Club: @NeuroImmunology [43]; MuscularDystrophyUK: @MDUK_News [44]; and Sign Against Stroke: @signagnststroke [45]) for enabling interactions between clinicians and patients, using tweet replies, and following particular handles or hashtags. This allows patients suffering from a particular illness to directly connect with their clinicians and get clarification on their illness and further treatment prospects. In the longer term, this primarily addresses patients' main concerns but also improves the clinical community's awareness of understanding of the disease and its management, thus bridging critical care gaps.

Pitfalls: Ethical and Legal Concerns

Overview

The legal and ethical issues surrounding Twitter are no less complex than those affecting the internet in general. Currently, Twitter takes some precautions in identifying the authenticity of individual users. If one chooses to make bad decisions with

XSL•FO
RenderX

the information they are sharing, such as defamatory, harassing, malicious, menacing, deceptive, impersonating, or threatening tweets, they may be held liable for legal action [46]. However, regulations remain porous. Many professional bodies have issued formal guidance, such as the General Medical Council ethical guidance in the United Kingdom, Twitter Guidelines & Best Practices by the US Centers for Disease Control and Prevention, Twitter Best Practices & Tips for Physicians by Johns Hopkins Medicine, and numerous others. Some critical concerns are discussed in the following sections.

Misrepresentation of Credentials

Individuals posing as physicians may post information that tends to be taken at face value by an unsuspecting user.

Misinformation

Misleading messages may be widely retweeted and, hence, gain traction. No structured system currently exists to filter out misinformation on Twitter, potentially allowing inaccurately informed persons, nonprofessionals, and those with commercial interests to frame their views as facts. This can pose serious problems for trainees and patients alike, who may struggle to identify genuine and credible posts. Various groups may use the same hashtag, which confuses hashtag followers. Individuals or organizations may use a hashtag to either spread misinformation or to "bog down" a hashtag with irrelevant or unhelpful posts, which makes it harder to find the desired posts. In the nascent stages of the COVID-19 pandemic, there was a "flood of misinformation" [47] on Twitter, such that the World Health Organization termed it a "massive infodemic" [48].

Shahi et al undertook an exploratory study on COVID-19 misinformation on Twitter. Analysis of 1500 tweets relating to 1274 false and 276 partially false claims revealed that verified Twitter handles, including organizations and celebrities, were also involved in either creating or spreading the misinformation via new tweets or retweets, respectively. They also noted that false claims propagated faster than partially false claims (χ^2_3=10.2; P<.001; N=1500) [49]. Controversial posts were taken down only after large-scale objections [2,50]. In October 2020, Twitter removed a tweet by Scott Atlas, a controversial US scientist, in which he had wrongly stated that masks fail to protect against coronavirus: "Masks work? NO" [51]. In the same month, then–US President Donald Trump tweeted that the United States had "learned to live with" flu season, "just like we are learning to live with Covid, in most populations far less lethal!!!" Twitter added a caveat the same message with a warning about "spreading misleading and potentially harmful information" [52].

Twitter is experimenting with adding brightly colored labels directly beneath lies or misinformation posted by famous users (Figure 1, F), in addition to a community reports feature where inaccurate and misleading information posted by public figures is corrected directly beneath a tweet by verified fact checkers and journalists. While this may benefit the medical community [53], it may be worthwhile to explore strategies that favor dissemination of evidence-based information.

Commercial Influences

Twitter posts with biased information may be promoted by parties with commercial interests. During the COVID-19 infodemic, the US Food and Drug Administration had to issue warnings to individuals to refrain from promoting or selling colloidal silver on social media profiles [54]. Twitter may be potentially misused by individuals and companies peddling unproven, investigational, and controversial medications, and by organizers of predatory conferences [55].

In June 2020, an Indian-based Ayurveda pharmacy claimed that it had discovered a potential cure for COVID-19, with an "80 per cent cure rate" [56]. A tweet from the chief executive officer of the pharmacy, apparently suggesting that it had obtained approval from the Ministry of AYUSH (Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homeopathy) for the drug as a "cure for COVID-19," went viral on Twitter [57-60]. The combined drug preparation and its individual components together recorded sales of over 85 lakh units within 4 months of its launch, garnering an estimated 241 crore Indian rupees (~US $32.5 million) for the pharmacy [61]. These record sales occurred despite the fact that soon after its launch, the Ministry of AYUSH ordered the Ayurveda pharmacy to stop advertising the product as a COVID-19 cure due to the absence of appropriate clinical trial data to back its claim [61].

Similarly, predatory conferences aggressively solicit conference abstract submissions [62] and combine broad topics from multiple disciplines to cast a larger net and bring more physicians into their fold [63]. Social media, including Twitter, allows them easy access to a broad audience. Cress et al collated a list of confirmed predatory conference organizers [63]. Individuals often share their experience of receiving invitations for predatory conferences on Twitter; for example, Wim Crusio, the editor-in-chief of the journal Behavioral and Brain Functions (@WimCrusio) once shared the following [64]:

> *Just got an invitation for "Neurology 2019", a world-wide meeting that attracted no less than 60 participants last year in Paris! This gem is organized by @Euroscicon. #Avoid #PredatoryConferences #FakeConferences.*

Strategies to avoid such commercial misuse may include identifying and calling out suspicious activity or insisting on *identifier* hashtags.

Exposure Bias

It is possible to be highly influenced by a small group of opinions. Users should take time to consider whether what they see is unbiased, as algorithms suggest potential profiles of interest and posts based on the interests users have already demonstrated. Twitter is an enviable educational resource, but users must always ensure that their practice is appropriately evidence based.

Patient Confidentiality Breaches and Privacy Concerns

Twitter provides easy access to potential *research participant* data sets, which raises some concerns. In one study, 60% of Twitter users surveyed were unaware that publicly available tweets can be used for research, and 65% felt researchers should

XSL•FO
RenderX

not be able to use public tweets without user permission [65-67]. There are numerous examples of patient-identifiable information being inadvertently or inappropriately shared by professionals [68].

Unprofessional Behavior

Previous studies on verified medical Twitter users have reported a worrying trend of "unprofessional" tweets, including patient privacy violations, profanity, sexually explicit content, discriminatory statements, inaccurate information, and self-promotion [69]. While most of the violations were clearly not in keeping with a professional code of conduct, the interpretation of unprofessional was variable. Recently, a paper called *Prevalence of unprofessional social media content among young vascular surgeons* was retracted by the Journal of Vascular Surgery [70]. The researchers classified "pictures in a bikini, posted on Twitter" as unprofessional, leading to a backlash from the scientific community that was centered upon interpretation and boundaries of unprofessionalism. While some professional bodies have published ethical guidelines for doctors' use of social media, the field remains largely unregulated [71,72]. Twitter users should keep in mind that individuals have been dismissed from their professional positions as their tweets were deemed as "unprofessional behavior" by their institutions [73].

Breach of Intellectual Property

Sharing of prepublished content, speaker slides, and pictures taken during lectures beg the question of whether this may qualify as a breach of intellectual property. For users, it may be best to ask for permission from the primary source prior to posting information on Twitter.

Trolling

Trolling refers to deliberate acts of making controversial comments in order to provoke internet users to respond emotionally. Trolls have used tweets to target patients with photosensitive epilepsy with flashing images intended to provoke seizures in the recent past [74,75]. Before posting, users should consider whether their posts might be considered provocative.

Confusing Hashtag Use

Various groups may use the same hashtag and confuse hashtag followers. People may also use a hashtag incorrectly or to spread false information. In addition, people may bog down a hashtag with many posts that are not necessarily relevant or helpful. All of these make it harder to find the desired posts.

Tips for Using Twitter

There are numerous other ethical concerns associated with physicians' use of Twitter. One critical question is, "Who will regulate its use and how could regulations be enforced?" While Twitter has initiated, and will initiate, changes based on real-life experiences, the onus of patient confidentiality and professionalism lies with individual users. The challenges can be much higher in countries with lower English literacy, since the vast majority of information in the aforementioned resources is presented in English. The issue of language translation should be considered, as the message could be lost or transformed by translation. With this in mind, we propose some tips for using Twitter in Textbox 1.

XSL•FO
RenderX

Textbox 1. Tips for Twitter use in order to maintain a healthy, informative, and interactive Twitter profile.

How should I use Twitter?:

1. Consider operating separate personal and professional accounts.
2. Make an honest, representative Twitter profile. Identities are periodically verified by Twitter.
3. Before sharing or retweeting, check the authenticity of the information and source. If you should reference another user's work or tweet, use the retweet function and add a comment or use the relevant hashtag or handle to credit them. Alternatively, provide a reference or a URL to the source material.
4. Healthy critique is positive. Even if offence is not intended, comments that could be interpreted as bullying or discriminatory are ill advised. Consider whether foul language is truly necessary.
5. Educate yourself with the social media guidelines of your institution and strictly adhere to them. Remember that employers often review social media accounts—never tweet anything that could affect your professional standing or employability. In case of complaints and discrepancies, employers, regulatory bodies, or legal agencies can become involved.
6. Use meaningful hashtags. Follow a large number of other accounts and follow your followers back to maximize the number of people who see your tweets.
7. Proudly share your work. If people learn from you, they may like, follow, or retweet you.
8. When following, posting, and retweeting, endeavor to show good judgment and consider why you are saying what you are—ensure it reflects well on you and your colleagues.
9. Carefully choose the people and groups you follow or share information with. Review the privacy settings to regulate who can view your posts.
10. If you are a victim of targeted harassment or abuse, unfollow and end any communication with that user. If the behavior continues, it is recommended to block the user and report them to Twitter.

How to maintain social media professionalism:

1. Avoid advertising.
2. Avoid sharing any patient-identifiable information; even if anonymized, ensure you have the proper consent.
3. Demonstrate good judgement—if you should not say it in your workplace, it should remain untweeted. Avoid posting and sharing misinformation, racist or discriminatory statements, or unpleasant or explicit content. Think carefully before tweeting: if necessary, compose your tweet and review it after a short period before posting.

Conclusions

The Twitterverse is a vast and exponentially growing expanse of information. Neurologists have taken to Twitter to educate, promote research, share information rapidly, and reach a broader potential global audience. Twitter has added a new dimension to learning and education in neurology in a practical and interactive manner. Multichannel interactions have furthered improvements in patient care and other health care–related activities. This has been particularly beneficial in promoting, furthering, and communicating research. In addition, there is enhanced support available to patients, particularly in the fields of rare and orphan diseases. However, the pros of the Twitterverse must be balanced with the potential risks, which are common to all social media platforms. Greater reach means that misinformation and promotion of vested interests can be too easily and widely disseminated. Unregulated use can result in inappropriate and unprofessional conduct whether intentional or not, highlighting the urgent need for designing and implementing institutional, national, or specialty-specific ethical and legal guidelines on appropriate social media use. Nonetheless, with judicious use, the benefits of Twitter for a neurologist or neuroscientist outweigh the risks; a graphical rendering is provided in Figure S1 in Multimedia Appendix 1.

Acknowledgments

The authors acknowledge Dr Lara J Cooke, Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, for her immense help in supervision of the work during the entire time period, writing assistance, language editing and proofreading, and providing intellectual content. RP is supported by a Medical Research Council Clinician Scientist Fellowship (MR/S002065/1). RP, PT, and VV are supported by a Medical Research Council strategic award to establish an International Centre for Genomic Medicine in Neuromuscular Diseases (MR/S005021/1). CD is supported by funding from the Muscular Dystrophy Association of the United States.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
A graphical rendering of pros and cons of using Twitter from a neurologist's perspective.
[PNG File , 562 KB - jmir_v23i5e25229_app1.png]

References

1. About. Twitter. URL: https://about.twitter.com/en_us.html [accessed 2020-08-07]
2. Scroll Staff. Twitter removes Donald Trump's retweeted claim downplaying coronavirus toll in US. Scroll. 2020 Sep 01. URL: https://scroll.in/latest/971935/twitter-removes-donald-trumps-retweeted-claim-downplaying-coronavirus-toll-in-us [accessed 2020-12-20]
3. MS Focus: The Multiple Sclerosis Foundation. Twitter. URL: https://twitter.com/ms_focus [accessed 2020-08-07]
4. FSHD Society. Twitter. URL: https://twitter.com/fshdsociety [accessed 2020-08-07]
5. Sedrak MS, Cohen RB, Merchant RM, Schapira MM. Cancer communication in the social media age. JAMA Oncol 2016 Jun 01;2(6):822-823. [doi: 10.1001/jamaoncol.2015.5475] [Medline: 26940041]
6. Bisset CN, Carter B, Law J, Hewitt J, Parmar K, Moug SJ, ELF Study Group Collaborative Authorship. The influence of social media on recruitment to surgical trials. BMC Med Res Methodol 2020 Jul 28;20(1):201 [FREE Full text] [doi: 10.1186/s12874-020-01072-1] [Medline: 32723388]
7. AVERT DOSE trial. Twitter. URL: https://twitter.com/avertdosetrial [accessed 2020-08-07]
8. NovartisOncTrialsUS status. Twitter. 2019 Jan 31. URL: https://twitter.com/NovartisOncCT/status/1091033390541877258 [accessed 2020-12-17]
9. Wasilewski MB, Stinson JN, Webster F, Cameron JI. Using Twitter to recruit participants for health research: An example from a caregiving study. Health Informatics J 2019 Dec;25(4):1485-1497 [FREE Full text] [doi: 10.1177/1460458218775158] [Medline: 29843545]
10. Callister MN, Robbins MS, Callister NR, Vargas BB. Tweeting the headache meetings: Cross-sectional analysis of Twitter activity surrounding American Headache Society conferences. Headache 2019 Apr;59(4):518-531. [doi: 10.1111/head.13500] [Medline: 30891749]
11. Charkhchi P, Sahraian S, Beheshtian E, Carlos RC, Yousem DM, Sair HI. Missed opportunity: Neuroradiology training programs and social media. AJR Am J Roentgenol 2019 May;212(5):1136-1141. [doi: 10.2214/ajr.18.20707]
12. Bhayani RK, Fick L, Dillman D, Jardine DA, Oxentenko AS, O'Glasser A. Twelve tips for utilizing residency program social media accounts for modified residency recruitment. MedEdPublish 2020 Aug 26;9(1):2-14. [doi: 10.15694/mep.2020.000178.1]
13. Zelikovich A, Lawson E, Robbins M. The role of Twitter in neurology residency recruitment in 2020 and beyond. Neurology Blogs. 2020 Oct 14. URL: https://blogs.neurology.org/rf/the-role-of-twitter-in-neurology-residency-recruitment-in-2020-and-beyond/ [accessed 2020-10-14]
14. Charidimou A, Shams S. Introducing @microbleeds: A pilot Twitter space for cerebral microbleeds research. Int J Stroke 2016 Apr;11(3):NP40-NP41. [doi: 10.1177/1747493016632251] [Medline: 26865156]
15. Lafferty NT, Manca A. Perspectives on social media in and as research: A synthetic review. Int Rev Psychiatry 2015 Apr;27(2):85-96. [doi: 10.3109/09540261.2015.1009419] [Medline: 25742363]
16. MND Association. Twitter. URL: https://twitter.com/mndassoc [accessed 2020-12-20]
17. Parkinson's Foundation. Twitter. URL: https://twitter.com/ParkinsonDotOrg [accessed 2020-12-20]
18. Peters ME, Uible E, Chisolm MS. A Twitter education: Why psychiatrists should tweet. Curr Psychiatry Rep 2015 Dec;17(12):94. [doi: 10.1007/s11920-015-0635-4] [Medline: 26463050]
19. Huber A. Twitter journal clubs: Sharing knowledge from a social distance. Scope. 2020 Mar 31. URL: https://scopeblog.stanford.edu/2020/03/31/twitter-journal-clubs-sharing-knowledge-from-a-social-distance/ [accessed 2020-12-16]
20. Mandavilli A. Peer review: Trial by Twitter. Nature 2011 Jan 20;469(7330):286-287. [doi: 10.1038/469286a] [Medline: 21248816]
21. Tracey Milligan, MD, MS, FAAN. Twitter. URL: https://twitter.com/tracey1milligan [accessed 2020-08-08]
22. Bernstein L. Tweetorials — From early beginnings to huge growth and beyond. Symplur. 2019 Dec 09. URL: https://www.symplur.com/blog/tweetorials-from-early-beginnings-to-huge-growth-and-beyond/ [accessed 2020-08-19]
23. Breu AC. Why is a cow? Curiosity, tweetorials, and the return to why. N Engl J Med 2019 Sep 19;381(12):1097-1098. [doi: 10.1056/NEJMp1906790] [Medline: 31532957]
24. Breu AC. From tweetstorm to tweetorials: Threaded tweets as a tool for medical education and knowledge dissemination. Semin Nephrol 2020 May;40(3):273-278. [doi: 10.1016/j.semnephrol.2020.04.005] [Medline: 32560775]
25. Tweetorial. Urban Dictionary. URL: https://www.urbandictionary.com/define.php?term=Tweetorial [accessed 2020-12-20]
26. Chary MA, Chai PR. Tweetchats, disseminating information, and sparking further scientific discussion with social media. J Med Toxicol 2020 Apr;16(2):109-111 [FREE Full text] [doi: 10.1007/s13181-020-00760-0] [Medline: 32096006]
27. #mindmoments [Twitter search]. Twitter. URL: https://twitter.com/search?q=%23mindmoments [accessed 2020-08-07]
28. Ford KC, Veletsianos G, Resta P. The structure and characteristics of #PhDchat, an emergent online social network. J Interact Media Educ 2014 Mar 28;2014(1):8. [doi: 10.5334/2014-08]

XSL•FO
RenderX

29. Rebecca E. Fasano, MD status. Twitter. 2020 Dec 17. URL: https://twitter.com/RebeccaFasanoMD/status/1339570620024180736 [accessed 2020-12-17]
30. The Notorious EEG (M. Scott Perry MD). Twitter. URL: https://twitter.com/TheNotoriousEEG [accessed 2020-12-21]
31. IFCN and its Clinical Neurophysiology journals. Twitter. URL: https://twitter.com/ClinicalNeuroph [accessed 2020-12-21]
32. MDedge Neurology. Twitter. URL: https://twitter.com/MDedgeNeurology [accessed 2020-12-25]
33. Busis NA, Shanafelt TD, Keran CM, Levin KH, Schwarz HB, Molano JR, et al. Burnout, career satisfaction, and well-being among US neurologists in 2016. Neurology 2017 Jan 25;88(8):797-808. [doi: 10.1212/wnl.0000000000003640]
34. Levin KH, Shanafelt TD, Keran CM, Busis NA, Foster LA, Molano JRV, et al. Burnout, career satisfaction, and well-being among US neurology residents and fellows in 2016. Neurology 2017 Jun 30;89(5):492-501. [doi: 10.1212/wnl.0000000000004135]
35. Compton MT, Frank E. Mental health concerns among Canadian physicians: Results from the 2007-2008 Canadian Physician Health Study. Compr Psychiatry 2011;52(5):542-547. [doi: 10.1016/j.comppsych.2010.10.002] [Medline: 21129737]
36. Ice Bucket Challenge. Twitter. URL: https://twitter.com/icebucketals [accessed 2020-08-07]
37. Dhawan E. Why did the ALS Ice Bucket Challenge go viral? Ellevate. URL: https://www.ellevatenetwork.com/articles/5765-why-did-the-als-ice-bucket-challenge-go-viral [accessed 2020-12-19]
38. Understanding the impact of the Ice Bucket Challenge on the ALS Association's finances. ALS Association. 2019 Jun 05. URL: https://www.als.org/stories-news/understanding-impact-ice-bucket-challenge-als-associations-finances [accessed 2020-12-14]
39. McNeil K, Brna PM, Gordon KE. Epilepsy in the Twitter era: A need to re-tweet the way we think about seizures. Epilepsy Behav 2012 Mar;23(2):127-130. [doi: 10.1016/j.yebeh.2011.10.020] [Medline: 22134096]
40. Myers L, Jones J, Boesten N, Lancman M. Psychogenic non-epileptic seizures (PNES) on the internet: Online representation of the disorder and frequency of search terms. Seizure 2016 Aug;40:114-122 [FREE Full text] [doi: 10.1016/j.seizure.2016.06.018] [Medline: 27394057]
41. Sirven JI. It's time for an epilepsy Twitter revolution. Epilepsy Behav 2012 Mar;23(2):97. [doi: 10.1016/j.yebeh.2011.11.007] [Medline: 22227033]
42. Meng Y, Elkaim L, Wang J, Liu J, Alotaibi NM, Ibrahim GM, et al. Social media in epilepsy: A quantitative and qualitative analysis. Epilepsy Behav 2017 Jun;71(Pt A):79-84. [doi: 10.1016/j.yebeh.2017.04.033] [Medline: 28554148]
43. NeuroImmunology Club. Twitter. URL: https://twitter.com/neuroimmunology [accessed 2020-08-07]
44. MuscularDystrophyUK. Twitter. URL: https://twitter.com/MDUK_News [accessed 2020-12-20]
45. Sign Against Stroke. Twitter. URL: https://twitter.com/signagnststroke [accessed 2020-12-20]
46. About rules and best practices with account behaviors. Twitter. URL: https://help.twitter.com/en/rules-and-policies/twitter-rules-and-best-practices [accessed 2020-08-11]
47. Twitter to label misinformation about coronavirus amid flood of false claims and conspiracy theories. CBS news. 2020 May 13. URL: https://www.cbsnews.com/news/twitter-misinformation-disputed-tweets-claims-coronavirus/ [accessed 2020-12-15]
48. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. 2020 Sep 23. URL: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation [accessed 2020-12-15]
49. Shahi G, Dirkson A, Majchrzak T. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021 Mar;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
50. Fleming N. Coronavirus misinformation, and how scientists can help to fight it. Nature 2020 Jul;583(7814):155-156. [doi: 10.1038/d41586-020-01834-3] [Medline: 32601491]
51. Pilkington E. Twitter removes false coronavirus tweet by Trump's favourite health adviser. The Guardian. 2020 Oct 19. URL: http://www.theguardian.com/us-news/2020/oct/19/scott-atlas-twitter-removal-false-coronavirus-tweet [accessed 2020-09-19]
52. Trump Covid post deleted by Facebook and hidden by Twitter. BBC News. 2020 Oct 06. URL: https://www.bbc.com/news/technology-54440662 [accessed 2020-08-07]
53. Collins B. Twitter is testing new ways to fight misinformation — including a community-based points system. NBC News. 2020 Feb 20. URL: https://www.nbcnews.com/tech/tech-news/twitter-testing-new-ways-fight-misinformation-including-community-based-points-n1139931 [accessed 2020-08-07]
54. Hauser C, Diaz J. FDA warns 7 companies to stop claiming silver and other products treat coronavirus. The New York Times. 2020 Mar 09. URL: https://www.nytimes.com/2020/03/09/health/fda-letter-coronavirus-cures.html [accessed 2020-12-16]
55. Kaestner V, Brown A, Tao D, Prasad V. Conflicts of interest in Twitter. Lancet Haematol 2017 Sep;4(9):e408-e409. [doi: 10.1016/S2352-3026(17)30109-6] [Medline: 28863800]
56. Patanjali claims they have 'cure' for COVID-19 in Ayurvedic 'Coronil' but why is Twitter outraging? NEWS18. 2020 Jun 14. URL: https://www.news18.com/news/buzz/

XSL•FO
RenderX

patanjali-claims-they-have-cure-for-covid-19-in-ayurvedic-coronil-but-why-is-twitter-outraging-2668985.html [accessed 2020-12-20]
57. Chattopadhyay A. Fact check: Misleading Tweet from Patanjali CEO goes viral as AYUSH Ministry's approval for Coronil. The Logical Indian. 2020 Jun 26. URL: https://thelogicalindian.com/fact-check/ministry-of-ayush-patanjali-coronil-21939 [accessed 2020-12-21]
58. Sutaria S. Never said Coronil can cure corona: Patanjali CEO makes false claim. BOOM. 2020 Jun 30. URL: https://www.boomlive.in/fake-news/never-said-coronil-can-cure-corona-patanjali-ceo-makes-false-claim-8700 [accessed 2020-12-21]
59. Patanjali Ayurved status. Twitter. 2020 Jun 23. URL: https://twitter.com/PypAyurved/status/1275342291234770944 [accessed 2020-12-21]
60. Tijarawala SK status. Twitter. 2020 Jun 23. URL: https://twitter.com/tijarawala/status/1275279384698482688 [accessed 2020-12-21]
61. Chandna H. ThePrint. 2020 Oct 27. URL: https://theprint.in/health/not-a-cure-for-covid-but-patanjalis-coronil-kit-drugs-are-a-big-hit-85-lakh-units-sold/531316/ [accessed 2020-12-20]
62. Cobey KD, de Costa e Silva M, Mazzarello S, Stober C, Hutton B, Moher D, et al. Is this conference for real? Navigating presumed predatory conference invitations. J Oncol Pract 2017 Jul;13(7):410-413. [doi: 10.1200/jop.2017.021469]
63. Cress P. Are predatory conferences the dark side of the open access movement? Aesthet Surg J 2017 Jun 01;37(6):734-738. [doi: 10.1093/asj/sjw247] [Medline: 28158556]
64. Wim Crusio status. Twitter. 2019 Jan 22. URL: https://twitter.com/WimCrusio/status/1087692265819123712 [accessed 2020-12-20]
65. Williams M, Burnap P, Sloan L, Jessop C, Lepps H. Users' views of ethics in social media research: Informed consent, anonymity, and harm. In: Woodfield K, editor. The Ethics of Online Research (Advances in Research Ethics and Integrity, Vol. 2). Bingley, UK: Emerald Publishing Limited; 2017:27-52.
66. Chiauzzi E, Wicks P. Digital trespass: Ethical and terms-of-use violations by researchers accessing data from an online patient community. J Med Internet Res 2019 Feb 21;21(2):e11985 [FREE Full text] [doi: 10.2196/11985] [Medline: 30789346]
67. Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth 2017;8(4):266-276. [doi: 10.1080/23294515.2017.1403980] [Medline: 29125425]
68. Leach B. 'Tweeting' breaches patient confidentiality. The Telegraph. 2009 Sep 24. URL: https://www.telegraph.co.uk/technology/twitter/6225315/Tweeting-breaches-patient-confidentiality.html [accessed 2020-08-19]
69. Chretien KC, Azar J, Kind T. Physicians on Twitter. JAMA 2011 Mar 09;305(6):566-568. [doi: 10.1001/jama.2011.68] [Medline: 21304081]
70. Hardouin S, Cheng TW, Mitchell EL, Raulli SJ, Jones DW, Siracuse JJ, et al. Prevalence of unprofessional social media content among young vascular surgeons. J Vasc Surg 2020 Aug;72(2):667-671. Retracted in: J Vasc Surg. 2020 Oct;72(4):1514. doi: 10.1016/j.jvs.2020.08.018. [doi: 10.1016/j.jvs.2019.10.069] [Medline: 31882313]
71. Liu HY, Beresin EV, Chisolm MS. Social media skills for professional development in psychiatry and medicine. Psychiatr Clin North Am 2019 Sep;42(3):483-492 [FREE Full text] [doi: 10.1016/j.psc.2019.05.004] [Medline: 31358127]
72. Farnan JM, Snyder Sulmasy L, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: Patient and public relationships: Policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627. [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
73. Christine Hamilton axed from charity over burka tweet. BBC News. 2018 Aug 12. URL: https://www.bbc.com/news/uk-45160466 [accessed 2020-12-20]
74. Shead S. People with epilepsy targeted in Twitter attack. BBC News. 2019 Dec 18. URL: https://www.bbc.com/news/technology-50835214 [accessed 2020-12-18]
75. Wolfe E, Ahmed S. A Twitter cyberattack on the Epilepsy Foundation posted strobing images that could trigger seizures. CNN. 2019 Dec 17. URL: https://edition.cnn.com/2019/12/17/tech/epilepsy-strobe-twitter-attack-trnd/index.html [accessed 2020-08-13]

Abbreviations

4HL: 4H leukodystrophy
ALS: amyotrophic lateral sclerosis
AYUSH: Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homeopathy
EEG: electroencephalogram
ELF: Emergency Laparotomy and Frailty
MDS: Movement Disorder Society
MND: motor neuron disease

XSL•FO
RenderX

PMSF: Phelan McDermid Syndrome Foundation

Edited by R Kukafka; submitted 23.10.20; peer-reviewed by J Liu, Y Graham; comments to author 10.12.20; revised version received 06.01.21; accepted 16.03.21; published 14.05.21.

Please cite as:
Mishra B, Saini M, Doherty CM, Pitceathly RDS, Rajan R, Siddiqi OK, Ramdharry G, Asranna A, Tomaselli PJ, Kermode AG, Bajwa JA, Garg D, Vishnu VY
Use of Twitter in Neurology: Boon or Bane?
J Med Internet Res 2021;23(5):e25229
URL: https://www.jmir.org/2021/5/e25229
doi:10.2196/25229
PMID:33988522

©Biswamohan Mishra, Monica Saini, Carolynne M Doherty, Robert D S Pitceathly, Roopa Rajan, Omar K Siddiqi, Gita Ramdharry, Ajay Asranna, Pedro Jose Tomaselli, Allan G Kermode, Jawad A Bajwa, Divyani Garg, Venugopalan Y Vishnu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.05.2021. 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

Mobile Apps to Improve Medication Adherence in Cardiovascular Disease: Systematic Review and Meta-analysis

Shahd Al-Arkee[1], PhD; Julie Mason[1], PhD; Deirdre A Lane[2], PhD; Larissa Fabritz[3], FRCP; Winnie Chua[3], PhD; M Sayeed Haque[4], PhD; Zahraa Jalal[1], PhD

[1]Institute of Clinical Sciences, University of Birmingham, Birmingham, United Kingdom

[2]Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool, United Kingdom

[3]Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, United Kingdom

[4]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Shahd Al-Arkee, PhD
Institute of Clinical Sciences
University of Birmingham
Edgbaston
Birmingham, B15 2TT
United Kingdom
Phone: 44 0121 414 4042
Email: s.k.k.al-arkee@pgr.bham.ac.uk

Abstract

Background: Adherence rates of preventative medication for cardiovascular disease (CVD) have been reported as 57%, and approximately 9% of all CVD events in Europe are attributable to poor medication adherence. Mobile health technologies, particularly mobile apps, have the potential to improve medication adherence and clinical outcomes.

Objective: The objective of this study is to assess the effects of mobile health care apps on medication adherence and health-related outcomes in patients with CVD. This study also evaluates apps' functionality and usability and the involvement of health care professionals in their use.

Methods: Electronic databases (MEDLINE [Ovid], PubMed Central, Cochrane Library, CINAHL Plus, PsycINFO [Ovid], Embase [Ovid], and Google Scholar) were searched for randomized controlled trials (RCTs) to investigate app-based interventions aimed at improving medication adherence in patients with CVD. RCTs published in English from inception to January 2020 were reviewed. The Cochrane risk of bias tool was used to assess the included studies. Meta-analysis was performed for clinical outcomes and medication adherence, with meta-regression analysis used to evaluate the impact of app intervention duration on medication adherence.

Results: This study included 16 RCTs published within the last 6 years. In total, 12 RCTs reported medication adherence as the primary outcome, which is the most commonly self-reported adherence. The duration of the interventions ranged from 1 to 12 months, and sample sizes ranged from 24 to 412. Medication adherence rates showed statistically significant improvements in 9 RCTs when compared with the control, and meta-analysis of the 6 RCTs reporting continuous data showed a significant overall effect in favor of the app intervention (mean difference 0.90, 95% CI 0.03-1.78) with a high statistical heterogeneity (I^2=93.32%). Moreover, 9 RCTs assessed clinical outcomes and reported an improvement in systolic blood pressure, diastolic blood pressure, total cholesterol, and low-density lipoprotein cholesterol levels in the intervention arm. Meta-analysis of these clinical outcomes from 6 RCTs favored app interventions, but none were significant. In the 7 trials evaluating app usability, all were found to be acceptable. There was a great variation in the app characteristics. A total of 10 RCTs involved health care professionals, mainly physicians and nurses, in the app-based interventions. The apps had mixed functionality: 2 used education, 7 delivered reminders, and 7 provided reminders in combination with educational support.

Conclusions: Apps tended to increase medication adherence, but interventions varied widely in design, content, and delivery. Apps have an acceptable degree of usability; yet the app characteristics conferring usability and effectiveness are ill-defined. Future large-scale studies should focus on identifying the essential active components of successful apps.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019121385; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=121385

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e24190) doi:10.2196/24190

KEYWORDS

mobile health care applications; medication adherence; cardiovascular disease; systematic review; mobile phone

Introduction

Cardiovascular diseases (CVDs) are responsible for almost one-third of all deaths worldwide, leading to an estimated 17.9 million deaths each year [1]. A long-term use of cardiovascular medications significantly reduces the risk of morbidity and mortality [2,3], but their full therapeutic potential cannot be achieved if patients are nonadherent [4]. Approximately 9% of all CVD events in Europe are attributed to poor medication adherence [5], with adherence rates of only 57% [6].

Developing interventions to tackle medication nonadherence is important for improving health outcomes. A recent network meta-analysis of many different interventions showed that those with a technology-based approach had a positive, but short-lived, effect on medication adherence [7]. The escalating inclusion of technology into everyday life has witnessed the introduction of mobile health (mHealth) interventions, such as mobile apps, to support patients and health care professionals (HCPs) in disease management [8,9]. These reviews were not confined to app-based interventions. Some of the wide-ranging interventions included apps, whereas other mHealth interventions, such as text messaging and emails, were common. Several systematic reviews have indicated that apps may play a role in improving medication adherence in patients with CVD. For example, one systematic review included smartphone app-based interventions to promote lifestyle and behavior changes, reporting them as effective at improving medication adherence and increasing physical activity behavior [10]. For secondary prevention in patients with cerebrovascular disease, another systematic review showed improved medication adherence, a better maintenance of blood pressure (BP) and lipids within target ranges, and decreased episodes of angina, transient ischemic attack, and stroke with mHealth interventions, several of which included apps [11]. In contrast, a systematic review of internet-based interventions, which included apps, improved dietary outcomes, quality of life (QoL), and physical activity but reported a lack of evidence for their effect on medication adherence [12].

Published evidence for the beneficial effects of apps on medication adherence is often lacking or inconclusive. This study evaluates the effectiveness of app-based interventions on medication adherence in patients with CVD. Furthermore, this study explores the effects of app-based interventions on health-related outcomes, the functionality and usability of apps for patients, and the involvement of HCPs in the delivery of the intervention.

Methods

Search Strategy and Study Selection

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [13]. The protocol was registered in the PROSPERO database (CRD42019121385) [14].

MEDLINE (Ovid), PubMed Central, Cochrane Library, CINAHL Plus, PsycINFO (Ovid), Embase (Ovid), and Google Scholar databases were searched from inception to January 2020 using a 3-domain search strategy to include terms related to CVD, apps, and medication adherence (the search strategy is presented in Multimedia Appendix 1). Studies were selected if they were randomized controlled trials (RCTs), if they were published in English, if they were for patients with CVD (eg, atrial fibrillation, coronary heart disease, heart failure, hypercholesterolemia, hypertension, myocardial infarction, and stroke), and if the intervention included an app to improve medication adherence. A 2-stage process was used to select studies for inclusion in this review. First, 1 author (SA) screened titles and abstracts for relevance and removed duplicate records. Where ambiguities arose from the screening process, 1 of 2 different authors (JM and ZJ) independently assessed the title and abstract for relevance. For the second stage of screening, 2 authors (SA and ZJ) independently reassessed the full-text studies matching the prespecified criteria for eligibility. Bibliographies of selected studies were hand searched for additional references.

Data Extraction and Quality Assessment

Data extraction was conducted using a standardized form developed specifically for this review. Extracted data included the characteristics of the study and details of the intervention strategy. Quality assessment was conducted independently by 2 authors (SA and ZJ), and consensus was reached through discussion. The risk of bias was guided by the Cochrane Collaboration tool for RCTs [15], with the bias broadly categorized as selection, performance, attrition, or other bias. Once categorized, each bias domain was further categorized as low, high, or unclear risk of bias. Agency for Healthcare Research and Quality standards [15] were then applied, and an overall summary was generated using Review Manager (RevMan. version 5.4, The Cochrane Collaboration) [16].

Data Synthesis and Statistical Analysis

The outcome data were extracted from each trial. The authors were contacted for raw data where follow-up points for individual trials were identified but outcomes not reported in the published manuscripts. Four main analyses were conducted: (1) a series of meta-analyses of intervention effects on medication adherence at different time points of intervention duration; (2) a univariable meta-regression analysis, regressing the app intervention across trials on intervention duration; (3) a meta-analysis of intervention effects on medication adherence across all included trials at the final time point of intervention duration; and (4) a series of meta-analyses of intervention effects on systolic blood pressure (SBP), diastolic blood pressure (DBP), total cholesterol (TC), and low-density lipoprotein cholesterol (LDL-C) levels at the third month of the intervention. For the meta-analyses, trials reporting continuous data, means, SD, and sample sizes were included. Where SE or CI were

XSL•FO
RenderX

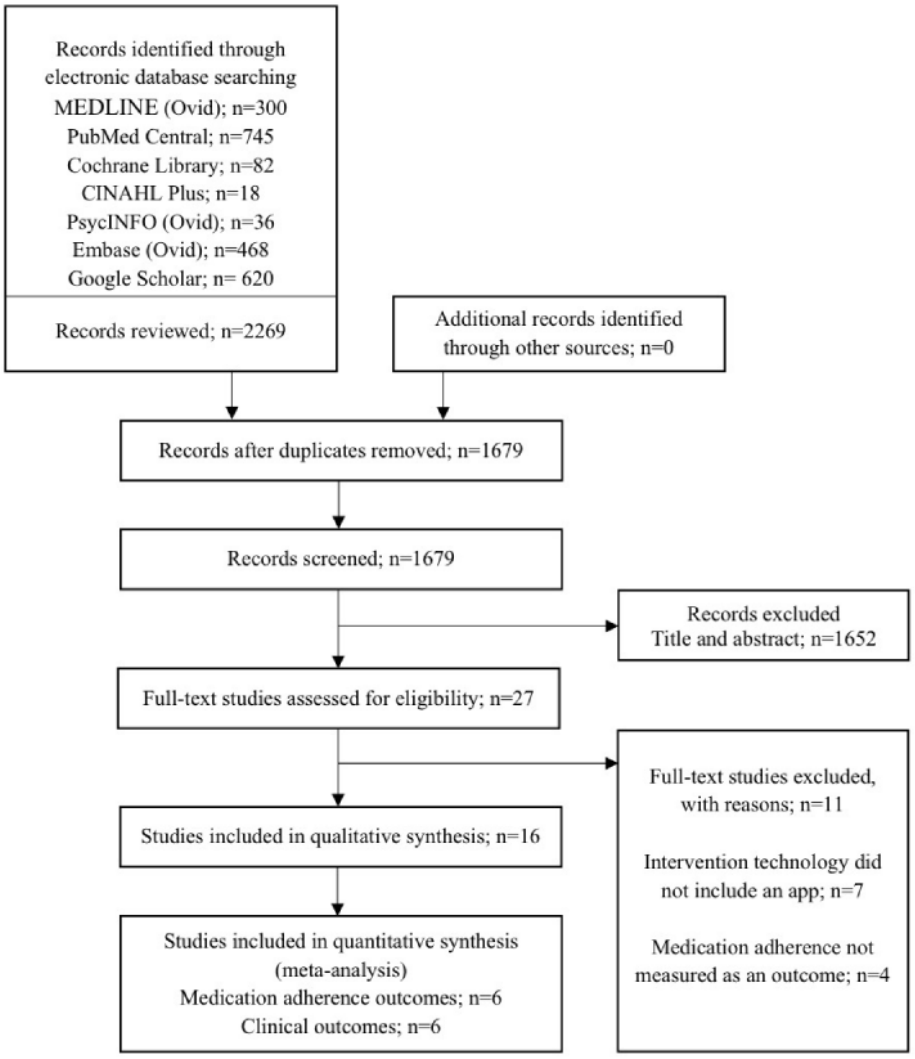
reported, the SD was manually calculated. A random-effects model was used to allow for differences in the true intervention effect across trials. The Q test was used to assess heterogeneity, with a significant result (P<.05) indicating heterogeneity across trials. The I^2 statistic was computed to describe the percentage of variability effect estimates due to heterogeneity. I^2 values of 25%, 50%, and 75% were assigned as low, moderate, and high heterogeneities, respectively [17]. The statistical package STATA (StataCorp, Stata Statistical Software: Release 16) was used for the meta-analysis [18].

Results

Search Results

Searches yielded 2269 citations, of which 590 duplicates were removed. The title and abstract screening resulted in 27 full-text review studies. Of these, 11 studies were excluded. No additional citations were identified by hand searching. Therefore, 16 RCTs were included in this review [19-34]. A PRISMA flowchart summarizing the study selection is shown in Figure 1.

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



Study Characteristics and Design

All the included studies were published between 2013 and 2019. They all compared 1 or more interventions (app alone or app in conjunction with a package of participant support) with a control arm described as usual care. A total of 10 studies randomized patients to parallel intervention or control groups [20,21,23,25-28,31-33]; 2 had a crossover design [19,30]; and 4 were cluster randomized by the trial site [22,24], researcher [29], or physician [34]. Study sample sizes ranged from 24 [30] to 412 [31], and interventions ranged in duration from 1 [23,30,32] to 12 [29] months. The definition of usual care in the control groups differed among the studies. It was defined as follow-up without the use of the app in 7 trials [22,24,26,27,29,31,33]; the use of the app with limited functionality in 1 trial [32]; and an alternative intervention not including apps, for example, the use of a SMS text message [20,28,34], follow-up phone calls [21], use of a pillbox [23], and use of an e-diary [25] in 6 trials. For the 2 crossover trials, nondigital technology methods were used [19,30] (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the included randomized controlled trials.

Source; country	RCT[a] design	Number of randomized participants	Intervention and control arm	Length of intervention	Primary and secondary outcomes measures
Brath et al [19]; Austria	Crossover 2-arm	77	• Intervention: monitoring phase using app • Control: control phase using paper diary	20 weeks	• Primary: medication adherence • Secondary: changes in SBP[b], DBP[c], HbA1c[d], LDL-C[e], and usability of the app
Chandler et al [20]; United States	Parallel 2-arm	54	• Intervention: SMASH[f] app • Control: SMS text messages on lifestyle tips unrelated to medication adherence	9 months	• Primary: change in SBP • Secondary: change in DBP, medication adherence, and patients' satisfaction with using the app
Fang and Li [21]; China	Parallel 3-arm	280	• Intervention: 2 arms: (1) SMS text messages using an app and (2) SMS text messages using an app plus micro letter • Control: phone	6 months	• Primary: medication adherence • Secondary: none
Frias et al [22]; United States	Parallel 3-arm, clustered by study site	118	• Intervention: a DMO[g] system designed to provide feedback about taking medication to both patients and providers consisted of an ingestible sensor, sensor patch, and app, 2 arms: (1) 4-week DMO and (2) 2.12-week DMO • Control: no system use	12 weeks	• Primary: change in SBP • Secondary: changes in SBP, DBP, HbA1c, and LDL-C; medication adherence; and satisfaction with using the app
Goldstein et al [23]; United States	Parallel 2×2 arm	60	• Intervention: 2 arms: (1) mHealth[h] app reminder and (2) mHealth app silent • Control: 2 arms: (1) telehealth (pillbox reminder) and (2) telehealth (pillbox silent)	28 days	• Primary: medication adherence • Secondary: acceptance of the app
Guo et al [24]; China	Parallel 2-arm, clustered by study site	209	• Intervention: mAF app • Control: no app use	3 months	• Primary: medication adherence, usability of the app, PAM[i], patients' knowledge, anticoagulation satisfaction, and QoL[j] • Secondary: none
Johnston et al [25]; Sweden	Parallel 2-arm	174	• Intervention: interactive patient support tool (web-based app) • Control: no app use, only simplified tool	6 months	• Primary: medication adherence • Secondary: change in SBP and LDL-C; QoL; and usability of the app
Kim et al [26]; Republic of Korea	Parallel 2-arm	95	• Intervention: Wireless Self-Monitoring, an app with enrolled in the HealthyCircles Platform • Control: no app use	6 months	• Primary: medication adherence, PAM, SBP, and DBP • Secondary: none
Labovitz et al [27]; United States	Parallel 2-arm	28	• Intervention: artificial intelligence app • Control: no daily monitoring	12 weeks	• Primary: medication adherence • Secondary: medication adherence for patients receiving DOACs[k] and usability of the app

XSL•FO
RenderX

Source; country	RCT[a] design	Number of randomized participants	Intervention and control arm	Length of intervention	Primary and secondary outcomes measures
Liu et al [28]; China	Parallel 2-arm	57	• Intervention: HeartGuardian app and weekly text messages on health education • Control: weekly SMS text messages on health education	12 weeks	• Primary: HDL-C[l], LDL-C, TC[m], and triglyceride • Secondary: medication adherence
Márquez Contreras et al [29]; Spain	Parallel 2-arm, clustered by researchers	154	• Intervention: AlerHTA app • Control: usual care in arterial hypertension	12 months	• Primary: medication adherence • Secondary: SBP and DBP
Mertens et al [30]; German	Crossover 2-arm	24	• Intervention: medication app • Control: a paper diary	28 days	• Primary: medication adherence • Secondary: user experience of the app
Morawski et al [31]; United States	Parallel 2-arm	412	• Intervention: Medisafe app • Control: no intervention	12 weeks	• Primary: medication adherence and change in SBP • Secondary: SBP and DBP
Ni et al [32]; China	Parallel 2-arm	50	• Intervention: BB app and WeChat app • Control: WeChat app	30 days	• Primary: medication adherence and heart rate, SBP, and DBP • Secondary: acceptability of the app
Santo et al [33]; Australia	Parallel 3-arm	166	• Intervention: 2 arms: (1) basic medication reminder app and (2) advanced medication reminder app • Control: standard care as determined by patients' physicians	3 months	• Primary: medication adherence • Secondary: BP[n], TC, LDL-C, and acceptability of the app
Sarfo et al [34]; Ghana	Parallel 2-arm, clustered by physician	60	• Intervention: Blue-toothed UA-767Plus BT BP device and a smartphone with an embedded app • Control: SMS text messages on healthy lifestyle behaviors	3 months	• Primary: BP • Secondary: medication adherence, hypertension management competence, autonomous self-regulation score for glucose control, patients' satisfaction with using the app, side effects of antihypertensive medications, hypertension, and stroke knowledge

[a]RCT: randomized controlled trial.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]HbA_{1c}: glycated hemoglobin.

[e]LDL-C: low-density lipoprotein cholesterol.

[f]SMASH: Smartphone Med Adherence Stops Hypertension.

[g]DMO: digital medicine offering system.

[h]mHealth: mobile health.

[i]PAM: patient activation measure.

[j]QoL: quality of life.

[k]DOAC: direct oral anticoagulant.

[l]HDL-C: high-density lipoprotein cholesterol.

[m]TC: total cholesterol.

[n]BP: blood pressure.

XSL•FO
RenderX

Participant Characteristics

The included trials covered a range of different CVDs and risk factors: atrial fibrillation [24], coronary heart disease [21,30,32,33], diabetes [19,22], heart failure [23], hypercholesterolemia [19], hypertension [19,20,22,26,29,31], myocardial infarction [25,28,30], and stroke [27,34]. The mean age of participants varied depending on the disease and ranged from 46.5 (SD 9.9) [20] to 73.8 (SD 7.5) years [30]. All studies recruited outpatients from secondary care [19,21,23-28,30,32,34], primary care [20,22,29], tertiary care [33], or web-based patient communities [31].

App Characteristics

The characteristics of the trialed apps are shown below. Each study used a different app developed by different organizations: 8 were academic or government institutions [20,21,23,24,26,28-30], whereas others were commercial organizations. A total of 7 apps were supported by platforms [19-22,24,26,27]. The functionality of the apps and platforms varied across the different trials and the interactions needed by patients. All but 2 apps [24,26] delivered medication reminders to promote medication adherence. For the majority, this was their primary function, with 4 apps using one-way SMS text message reminders [21,27,28,32] and 5 delivering a mobile device alert [23,29-31,33]. Others had a primary focus on self-monitoring alone [26] or with a medication reminder [25], patient education [24], or delivery of a tailored motivational SMS text message based on medication adherence levels [20,34]. Two trials used the app to transmit patients' adherence data to the associated platform to be monitored by HCPs [19,22].

Involvement of HCPs

Half of the trials involved physicians and/or nurses in app use [19,21,24,25,29,30,32,34] (Table 2). A trial involved pharmacists in blinding study medication, whereas the health care team, whose professions remained unspecified, used the app and monitored the data transmitted to the associated platform [22]. One trial permitted the sharing of patients' data with families and caregivers as well as with HCPs [26]. A total of 6 trials did not specify the type of HCPs involved in app use [20,23,27,28,31,33]. The involvement of HCPs varied; most of the trials involved HCPs to monitor patients' data [19,22,24-26,30,34], instruct patients on how to use the app [29], and send educational materials to patients via the app [21,32].

XSL•FO
RenderX

Table 2. Mobile app characteristics in the included randomized controlled trials.

Source	App name and functionality	Platform used with the app and functionality	Overall system functionality	Involvement of HCP[a]
Brath et al [19]	Name not specified, referred to as a mobile phone–based data gateway. Reader and transmitter of data from electronic medication blister to a remote database	Remote telemonitoring service: data sent from the app to platform and then analyzed for timing and number of pills taken, and an automatic reminder is sent to patients via SMS text messages	Reminder	Physician
Chandler et al [20]	SMASH[b] app: medication reminders via signals (blinking light, intermittent chime, automated SMS text messages, or phone call) and BP[c] monitor reminders via SMS text messages. The app provided timely tailored motivational and reinforcement SMS text messages based on the levels of medication adherence and SMS text message reminders to monitor BP with a Bluetooth-enabled BP device. The app also provided a cumulative table of average BP displayed in categories of daily, weekly, and/or monthly progress reports	HIPAA[d]-compliant servers: BP data sent from the app to platform, then analyzed for processing with timestamps, providing information for the calculation of adherence levels to the BP protocol	Reminder	Not stated
Fang and Li [21]	Name not specified, referred to as a messaging app: medication reminders via an SMS text messaging app, educational materials via micro letter	Huaxi-gold card: the platform sent SMS text messages, images, media content related to disease and other information at regular intervals	Reminder and education	Physician and nurse
Frias et al [22]	Proteus Discover app: reader and transmitter of the patient's adherence data from patch to the cloud and prompted the patient to take their medication doses as scheduled. Patients could visualize their data on their mobile devices via the app	Provider web portal: provider views summaries of the DMO[e] data for the patients on the web portal	Reminder and education	Clinic staff, pharmacist had a role in set up (coencapsulation of ingestible sensor and medication)
Goldstein et al [23]	Name not specified, referred to as a medication adherence app. Medication reminders provided via alert, patients could view list of medications with instructions, and they were able to record taking their medication	No platform	Reminder and education	Not stated
Guo et al [24]	mAF app: educational app used by both patients and physicians: For patients, personal health record (CHA_2DS_2-VASc[f], HAS-BLED[g], and SAMe-TT_2R_2[h] scores), patient educational programs (knowledge of atrial fibrillation and learn how to manage themselves at home), patient involvement in self-care items (monitor their heart rate, BP, and their quality of life), and structured follow-up consultation via a sent alert reminder. For physicians, clinical decision support	Cloud platform: data management	Education	Physician
Johnston et al [25]	Name not specified, referred to as an interactive patient support tool app: medication reminders via SMS text messages (e-diary) to register daily ticagrelor intake. Secondary prevention educational modules (exercise module, BMI module, and BP module)	No platform	Reminder and education	Physician and nurse
Kim et al [26]	HealthyCircles: an educational app that allowed patients and nurses to access the patient's reading recorded on the BP monitor devices. The BP measurements are wirelessly uploaded from BP devices to the HealthyCircles account	HealthyCircles platform: the platform sent reminders for self-monitoring BP, education information about the disease condition, and general health behavior recommendations	Education	Families, caregivers, and HCPs (profession not specified)

XSL•FO
RenderX

Source	App name and functionality	Platform used with the app and functionality	Overall system functionality	Involvement of HCP[a]
Labovitz et al [27]	Artificial intelligence app: medication reminders and dosing instructions via SMS text messages. Late doses generated notifications within the hour and before the end of the dosing window	Artificial intelligence platform: the platform sent an automatic SMS text message or emails to clinical staff if doses were missed, late, or based on incorrect use	Reminder	Clinic staff (profession not specified)
Liu et al [28]	HeartGuardian app: medication reminders via SMS text messages. The app provided educational materials; medication recording and daily feedback; and self-empowerment via automatic intelligent, real-time video feedback based on the subjects' medication adherence	No platform	Reminder and education	Not stated
Márquez Contreras et al [29]	AlerHTA app: medication and appointments reminders via alerts. The app recorded patients' personal data, the physician's advice about the prescribed treatment, and the results of the BP measurement. The app recommended BP levels as objectives	No platform	Reminder and education	Physician
Mertens et al [30]	iNephro medication plan app: medication reminders via alert, to support the drug intake needs of patients with chronic conditions on polypharmacy	No platform	Reminder	Physician
Morawski et al [31]	Medisafe app: medication reminders via alert. The app provided alerts to remind patients when it is time to take medications and generate weekly adherence reports, the app also allowed for tracking of BP and other biometric measurements	No platform	Reminder	Not stated
Ni et al [32]	BB reminder app and WeChat app: medication reminders via SMS text messages through the BB reminder app. Educational materials through the WeChat app	No platform	Reminder and education	Physician and nurse
Santo et al [33]	No specified name. Referred to as a medication reminder app. Medication reminders provided via alert. In the basic app, the reminders were noninteractive and occurred 1 time only, whereas the advanced app provided interactive and customizable features including daily reminders, which could be snoozed, rescheduled, and/or marked as a taken or missed dose; medication refill reminders; adherence statistics; and ability to share information with others such as family members, if the patient missed a medication dose	No platform	Reminder	Not stated

XSL•FO
RenderX

Source	App name and functionality	Platform used with the app and functionality	Overall system functionality	Involvement of HCP[a]
Sarfo et al [34]	No specified name. Referred to as medical regimen assistance app. Medication reminders provided via SMS text messages. The app reported BP measurements and medication intake and sent written and oral information on adherence criteria to take the medications within 2 hours of designated times and to measure BP every 3 days in the morning and evening	No platform	Reminder	Nurse

[a]HCP: health care professional.

[b]SMASH: Smartphone Medication Adherence Stops Hypertension.

[c]BP: blood pressure.

[d]HIPAA: Health Insurance Portability and Accountability.

[e]DMO: digital medicine offering system.

[f]CHA_2DS_2-VASc: congestive heart failure, hypertension, age>75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.

[g]HAS-BLED: Hypertension, abnormal renal or liver function, stroke, bleeding history or predisposition, labile international normalized ratio, age>65 years, drugs or alcohol concomitantly.

[h]SAMe-TT_2R_2: sex, age, medical history, treatment, tobacco use, and race.

Assessment of Medication Adherence

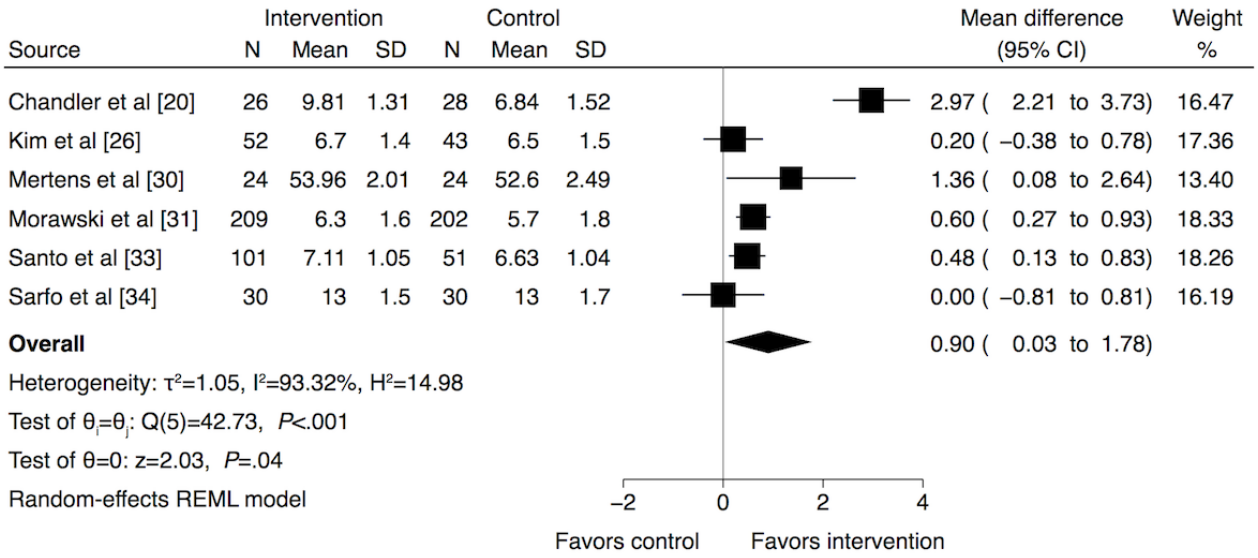
Adherence measures varied among studies (Multimedia Appendix 2 [19-34]). Most used questionnaires to include the validated 8-item Morisky Medication Adherence Scale [20,26,31,33] and the 4-item Morisky Medication Adherence Scale [21,28] and nonvalidated self-report questionnaires [24,25,30,32]. Other adherence measures included medication event monitoring systems (MEMSs) [29] and a digital medicine offering with an ingestible sensor taken alongside medication [22]. Other trials used a combination of measures; 2 trials combined 2 different measures, one for each arm. One trial used the remote medication adherence measurement system for the intervention and pill count for the control [19]. Another trial used an electronic self-report for the intervention and pillbox openings for the control [23]. Only 2 trials combined 2 different measures for both arms (pill counts and plasma samples [27] and pill counts and 8-item Morisky Medication Adherence Scale [34]).

Effect on Medication Adherence

Overall, 12 trials reported apps to enhance medication adherence rates [20,21,24,25,27-34], with 9 demonstrating significant improvement [20,24,25,28-31,33,34]. In the remaining 4 trials, 3 did not find a significant difference [22,23,26] and 1 reported a significant difference, with only 1 of the 4 medicines being monitored [19] (Multimedia Appendix 2 [19-34]). Six trials reporting continuous data were included in the meta-analysis of medication adherence [20,26,30,31,33,34]. Trials with the same duration of follow-up for the intervention were subjected to a separate meta-analysis and all favored the intervention, mean difference for month 1, 1.52 (95% CI 0.89 to 2.15); 2 trials [20,30] for month 3, 0.46 (95% CI 0.21 to 0.71); 4 trials [20,31,33,34], for month 6, 1.46 (95% CI −1.02 to 3.95); 2 trials [20,26], and for month 9, 1.49 (95% CI −1.42 to 4.40); 2 trials [20,34]. Meta-regression analysis for these 6 studies showed that the duration of intervention (ie, the follow-up month) did not exert a statistically significant impact on the effect of the app on medication adherence (P=.65). Thus, a combined meta-analysis (Figure 2) over the different trial durations was performed, thereby demonstrating a significant effect in favor of the app intervention (mean difference 0.90, 95% CI 0.03 to 1.78) with a high statistical heterogeneity (I^2=93.32%).

XSL•FO
RenderX

Figure 2. Meta-analysis results and forest plot of the effect of app-based interventions on medication adherence. Mean difference (95% CIs) are denoted by black boxes (black lines). The combined mean difference estimate for all studies is represented by a black diamond, where diamond width corresponds to 95% CI bounds. REML: restricted maximum likelihood.



Effect on Other Nonclinical Outcomes

An array of nonclinical outcomes was measured across the trials. Two trials have reported patient activation measures (PAMs) [22,26]. One trial reported a higher increase in PAM scores mean change for the intervention arm 7.9 (SE 2.4) when compared with control 1.7 (SE 3.3); mean difference 6.2 (SE 4.6), (95% CI −2.8 to 15.2) [22]. However, for the other trial, there was no significant difference in the average PAM score over the trial period (baseline: 78.0; end of trial: 76.0; *P*=.34) [26]. Patients' knowledge of their CVD was only reported in 2 trials [24,34], despite 9 of the 16 trials involving apps with an educational function [21-26,28,29,32]. Interestingly, one trial showed not only a significant improvement in knowledge with app use but also in medication adherence [24]. In the same trial, the benefits and burden of anticoagulation therapy were explored using a patient satisfaction questionnaire. Patients using the app expressed more anticoagulant *benefits*, whereas the control declared more *burden*: benefit (intervention: mean 15.6, SD 2.73 vs control: mean 14.21, SD 3.37; *P*=.05) and burden (intervention: mean 15.57, SD 6.57 vs control: mean 19.30, SD 6.39; *P*=.008) [24]. In the other trial, the knowledge questionnaire scores increased at the end of the trial but not significantly (intervention: mean 10.8, SD 0.8 vs control: mean 11.1, SD 1.1; *P*=.23) [34]. The QoL was assessed in 2 trials using the European Quality of Life–5 Dimensions measure [24,25]. One trial reported significantly higher QoL in the intervention arm compared with the control (*P*<.05; exact *P* value not quoted in original paper) [24], whereas in the other trial, QoL scores increased with app use over the duration of the trial but not significantly (*P*=.06) [25].

Effect on Clinical Outcomes

Clinical outcomes measured included BP, blood cholesterol, and blood glucose (Multimedia Appendix 3 [19,20,22,25,26,28,29,31-34]). Eight trials reported positive effects of apps on both SBP and DBP [19,20,22,25,26,29,33,34], and 4 reported significant results [19,20,22,29]. In total, 4 trials reported improvements in TC [19,22,28,33] and 3 were significant [19,22,28]. A reduction in LDL-C was observed with app-based interventions [19,22,25,28,33], but it was only significant in 2 trials [22,25]. Only 2 trials reported glycated hemoglobin (HbA_{1c}) as an outcome, with no significant change [19,22]. Meta-analysis for clinical outcomes was only possible at 3 months duration of intervention for SBP, DBP, TC, and LDL-C; all favored the use of an app in disease management, but not all were significant (Multimedia Appendix 4 [22,25,28,31,33,34]). Meta-analysis for HbA_{1c} was not possible because of the lack of reported outcomes.

App Usability, Acceptability, and Patient Satisfaction

Various questionnaires were used to evaluate the app usability for patients, but this was only done in 4 trials [19,24,25,27]. One study used a validated System Usability Scale to demonstrate greater usability in the app intervention arm than in the control arm (intervention: mean 87.3, SD 13.9 vs control: mean 78.1, SD 18.9; *P*=.001) [25]. Three trials evaluated app usability with nonvalidated questionnaires and obtained positive feedback from 80% or more of the participants [19,24,27]. Patients with stroke rated the app *extremely good* as a medication management tool and as means to improve physician-patient rapport [27]. Patients with atrial fibrillation agreed that the study app was user-friendly and helpful with additional positive feedback from physicians [24].

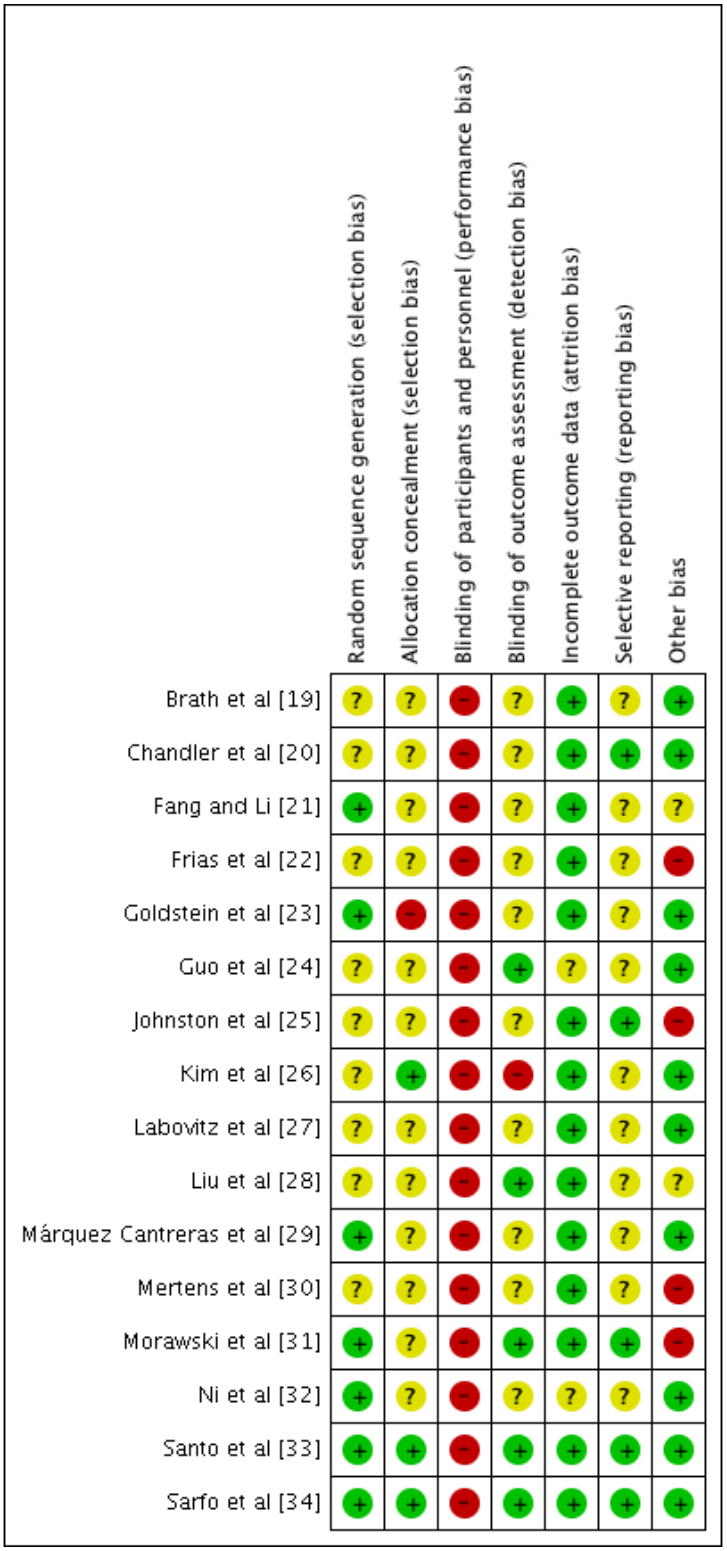
Four different trials explored app acceptability in patients [23,30,32,33]. Acceptance rates measured by nonvalidated questionnaires found the app to be more acceptable than the control [23], and most patients reported that the app was useful and helpful [33]. Interviews conducted within 2 studies revealed that patients accepted and appreciated receiving reminders and educational materials via the app [32] and that most patients (22/24) reported wanting to use the app in everyday life [30]. Three trials evaluated patient satisfaction with the apps being trialed by nonvalidated questionnaires, with more than 90% reporting the app as easy to use [20,22,34].

XSL•FO
RenderX

Risk of Bias of Included Trials

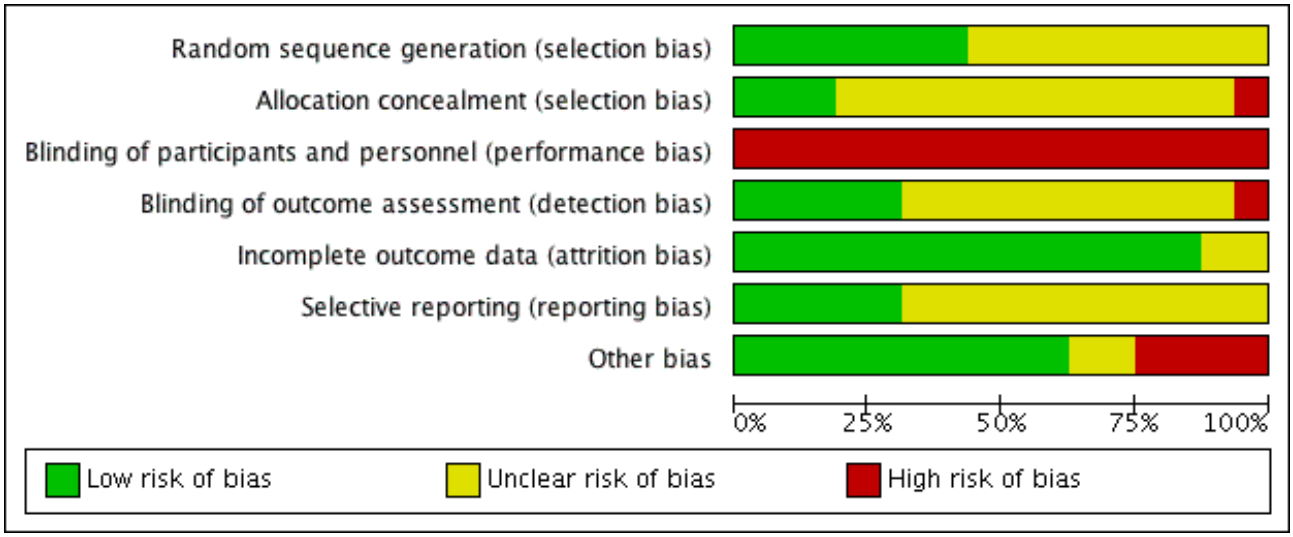
Only 7 trials reported sufficient random sequence generation [21,23,29,31-34], and only 3 trials reported allocation concealment [26,33,34]. Although these types of interventions are problematic to blind, outcome assessors could have been blinded, but only 5 trials clearly stated that this was done [24,28,31,33,34]. In total, 14 trials had a low risk of incomplete outcome data [19-23,25-31,33,34], whereas only 5 had a low risk of selective outcome reporting [20,25,31,33,34]. Ten trials had no other sources of risk of bias [19,20,23,24,26,27,29,32-34]. According to the Agency for Healthcare Research and Quality standards, most trials were considered to be of poor quality [19-32], with only 2 rated as fair [33,34]. Figures 3 and 4 present the risk of bias judgment.

Figure 3. Authors' judgments about each risk of bias item for each included trial. Green: low risk of bias; yellow: unclear risk of bias; red: high risk of bias.



XSL•FO
RenderX

Figure 4. Authors' judgments about each risk of bias item presented as percentages across all included trials.



Discussion

Principal Findings

This study included 16 RCTs that assessed the effectiveness of mobile app–based interventions on medication adherence [19-34]. A total of 9 trials showed a statistically significant improvement in medication adherence in the intervention arm [20,24,25,28-31,33,34]. The meta-analysis of 6 trials revealed that app interventions exert a significant positive effect on medication adherence with meta-regression, showing no statistically significant impact for the duration of use over a maximum of 9 months. However, the statistical and methodological heterogeneity was high [20, 26, 30, 31, 33, 34]. Ten trials assessed health-related outcomes and generally reported an improvement with intervention over control [19,20,22,24-26,28,29,33,34]. The apps used had mixed functionality, including reminders [19,20,27,30,31,33,34], education [24,26], or both [21-23,25,28,29,32]. Regarding the involvement of HCPs, most involved physicians and/or nurses [19,21,24,25,29,30,32,34]. The usability of apps was mainly assessed with questionnaires, with most participants reporting acceptance and ease of use [19,24,25,27]. The effectiveness of app interventions could not be assigned to particular app components or characteristics. Half of the trials were small-scale studies, that is, pilot studies [19,22,24,34] and feasibility studies [21,23,27,32], and most trials were classified as having poor quality of evidence because of the high risk of bias or insufficient reporting of information [19-32].

Relationship With Previous Published Literature

Previous systematic reviews have assessed the effectiveness of health care apps in the management of several different long-term conditions, including asthma [35], obesity and diabetes [36], and CVD [37]. Most included small-scale studies, with insufficient or low-quality evidence to support app use. Despite this, many reviews have reported beneficial trends, for example, in the promotion of positive behavior changes such as medication adherence [37].

A network meta-analysis of different interventions showed that technology-based interventions exert a major effect on the long-term management of medication adherence in patients with CVD [38]. The World Health Organization categorizes medication adherence measurements as either subjective or objective [39]. More than half of the trials in this systematic review used subjective self-report questionnaires to measure medication adherence [20,21,24-26,28,30-33], with a potential to overestimate adherence. Although there is no gold standard measure of medication adherence, a multi-measure approach is highly recommended to reduce subjectivity [40]. Therefore, the results of improved adherence from the trials included in this review should be interpreted with caution.

This review shows that objective measures can be improved with expected app use. For example, some of the trials included in this review assessed BP and showed improvements for participants in the intervention arms [19,20,22,25,26,29,33,34], a similar result to a previous systematic review assessing the effects of mobile apps designed for BP management [9]. Another systematic review and meta-analysis of 21 RCTs showed a reduction in HbA_{1c} levels in patients with diabetes [41]. In this review, the effectiveness of apps to support patients with diabetes was inconclusive, as only 2 included trials evaluated HbA_{1c}, and both the trials reported no significant difference in the change in HbA_{1c} between the intervention and control arms [19,22].

Although few trials included in this review investigated nonclinical outcomes other than adherence, those that did demonstrated a meaningful, but not always significant, improvement in PAM [22,26], patients' disease knowledge [24,34], anticoagulation satisfaction [24], and QoL [24,25]. These results align with existing systematic reviews of smartphone-based health care technologies, which demonstrate that apps could play an important role in patient education, self-management, and remote monitoring [42] and improvements in patients' QoL [37]. Furthermore, 2 pilot studies examining the feasibility of app use to enhance safe anticoagulation therapy and knowledge acquisition by patients showed a significant increase in anticoagulation knowledge after 3 months of app use [43,44]. The beneficial effects of apps on medication adherence will likely depend on the nature of the support needed

XSL•FO
RenderX

by different patients. To improve medication adherence, the literature suggests that some patients may need only reminders, whereas others need a greater knowledge and understanding of their disease and the medication prescribed [45,46]. There is a long history of reminders and patient education to improve medication adherence, and the introduction of app technology has seen these strategies incorporated into mHealth interventions. In this review, most of the included trials used apps with mixed functionality, including reminders, education, or both. All but 2 of the apps [24,26] included reminders [19-23,25,27-34]; of these, significant improvements in medication adherence were only reported in about half of the trials [20,24,25,28-31,33,34]. Thus, it remains impossible to assign success to a single component within a multifunctional intervention.

App design, user interface, and evaluation of these factors are often under-reported. In this review, 4 trials that assessed app usability demonstrated that the apps were user-friendly, and users were interested and engaged with the technology [19,24,25,27]. Three of these studies featured commercially developed apps [19,25,27]. The measures of app success for developers of commercial, academic, or government origins may explain why only 1 app developed by an academic institution [24] investigated usability. A systematic review of app usability in patients with diabetes also reported moderate to good usability, but users expressed preference for apps developed for tablet computers rather than smartphones due to their larger display and better illustrations [47]. Usability is a key factor in the uptake of mHealth apps [48,49], and it would make sense to conclude that a more user-friendly app might be more effective. In this review, significant improvements in medication adherence rates were found in only 2 of the 4 trials reporting a good usability [24,25]. This may, in part, be because usability outcomes measure ease of use (ie, user-friendliness) rather than motivation, engagement, and continued use. Motivating components, such as social contracts with family members and gamification, have been incorporated into some apps to improve their effectiveness [43]. Several studies highlight the importance of using theory to develop and design behavioral change interventions [50-52], which should also be considered in mHealth app intervention design. Only 3 of the trials in this review [20,28,34], reported the use of behavioral change theories to inform their app intervention, and it is of note that only one of the app interventions purported to involve social support or interactions outside of HCPs [26]. This review revealed that HCPs' involvement in app interventions for CVD health care mainly involved physicians and nurses [19,21,24,25,28-30,32,34], with 1 trial reporting pharmacist involvement; however, that did not include the administration of the app intervention [22]. With the widening clinical patient-facing roles of pharmacists within primary care [53,54] and reports of their effectiveness in both CVD management [55-57] and successful efforts to improve CVD medication adherence [58,59], it is potentially surprising that pharmacists were not more involved in any of these studies. The involvement of any HCP in the administration and concomitant use of apps with patients requires careful consideration. Such apps have the potential to increase HCP workload, and it remains unclear whether the cost of that involvement outweighs the benefits observed. Of the RCTs included in this review, 5 of the 9 that included HCPs in the administration of the app reported significant improvements in medication adherence, but no cost-benefit analysis was conducted [24,25,29,30,34]. In the current climate, with a growing choice of apps, a more important role for HCPs may be in the recommendation of safe, user-friendly, and effective mHealth apps for patients depending on their disease and apps chosen specifically to meet their patients' needs and motivations.

Strengths and Limitations

This review did not consider the differences in adherence between the medications included in the trials. Some medications might have a higher rate of nonadherence than others because of the adverse effects and taste of the formulation. The heterogeneity of the trials' methodologies, apps, and outcome measures studied made quantitative comparisons problematic. Different measures of medication adherence were used among the trials, which made it impossible to calculate the exact adherence rates. For several of the included trials, control groups were also subjected to an intervention aimed at improving medication adherence, meaning that the impact of the app intervention was not comparable with standard care. This, coupled with the potential for wide variations in standard care more generally, suggests that the findings of many of the included studies need to be interpreted with caution. Finally, this review included only RCTs; thus, other relevant studies and reports from the gray literature were excluded. However, RCTs are considered the cornerstone of clinical research to determine the efficacy of interventions and the highest level of evidence.

Implication for Practice and Policy

Health care apps have the potential to enhance medication adherence, leading to improvements in clinical and nonclinical outcomes in patients with CVD. However, the use of this technology to support medication adherence is in its infancy, and apps require robust testing to demonstrate its effectiveness. The trials included in this review provided inconsistent data regarding their effectiveness. Overall, user engagement and usability were rated positively, demonstrating interest in the concept. However, it is difficult to make strong, unrestricted recommendations for practice, especially with the methodological limitations of the included trials.

Implication for Research

This review indicates the need for further large-scale studies to determine whether mobile apps are effective in improving medication adherence among patients with CVD. There is a paucity of data to differentiate the effects of individual app intervention characteristics on behavioral change, and the most effective app functionality remains unknown. The involvement of HCPs in the use of mobile apps needs to be investigated further, needs to undergo cost-benefit analysis, and needs to be compared with the effectiveness standalone apps that do not require HCP input. Finally, a standard validated approach for medication adherence measurement is recommended for future studies to enable the comparison of findings and/or pooling of adherence data.

XSL•FO
RenderX

Conclusions

Mobile apps appear to enhance medication adherence and improve health-related outcomes. Apps have an acceptable degree of usability; yet the app characteristics conferring usability and effectiveness are often indeterminate due to their multifactorial design. Existing evidence is currently insufficient to unreservedly recommend the use of health care apps to improve adherence to CVD medications because of the generally small sample sizes; clinical and methodological heterogeneity between studies; and disparity in app features, content, and delivery, but they may enhance medication adherence as part of a package of care.

Authors' Contributions

SA, JM, and ZJ conceived and designed the study. SA, JM, and ZJ completed the screening and selection of studies, data extraction, and critical appraisal. SA, JM, and ZJ initially analyzed the data, with review and further interpretation from all other authors. SA, JM, and MSH performed the statistical data analysis. SA, JM, and ZJ drafted the manuscript. All authors contributed the manuscript review, revision, and final approval.

Conflicts of Interest

DAL is a coauthor of the mAF app but reports no other conflicts of interest in relation to this manuscript. LF has received institutional research grants from the European Union (for the design of mobile health apps for patient education), the British Heart Foundation, Medical Research Council (UK), and German Research Foundation. The Institute of Cardiovascular Research, University of Birmingham, has received an Accelerator Award by the British Heart Foundation (AA/18/2/34218) during the conduct of the study. LF is listed as an inventor on 2 patents held by the University of Birmingham (Atrial Fibrillation Therapy [WO 2015140571] and Markers for Atrial Fibrillation [WO 2016012783], not related to medical adherence or apps). LF has received grants from EU Horizon 2020 (CATCH ME Characterizing Atrial fibrillation by Translating its Causes into Health Modifiers in the Elderly (633196), and MAESTRIA Machine Learning Artificial Intelligence Early Detection Stroke Atrial Fibrillation [(965286) to fund WC). No other disclosures were reported.

Multimedia Appendix 1
Search strategy: Ovid MEDLINE (1946 to January 2020).
[PDF File (Adobe PDF File), 62 KB - jmir_v23i5e24190_app1.pdf]

Multimedia Appendix 2
Medication adherence of the included randomized controlled trials.
[PDF File (Adobe PDF File), 96 KB - jmir_v23i5e24190_app2.pdf]

Multimedia Appendix 3
Clinical outcomes of the included randomized controlled trials.
[PDF File (Adobe PDF File), 97 KB - jmir_v23i5e24190_app3.pdf]

Multimedia Appendix 4
Meta-analysis and forest plot of the effect of app-based interventions on clinical outcomes.
[PDF File (Adobe PDF File), 955 KB - jmir_v23i5e24190_fig.pdf]

References

1. Cardiovascular diseases (CVDs). World Health Organization. 2017. URL: http://www.who.int/mediacentre/factsheets/fs317/en/ [accessed 2020-04-15]
2. Data and statistics. World Health Organization. 2010. URL: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/data-and-statistics [accessed 2020-04-15]
3. Yusuf S. Two decades of progress in preventing vascular disease. Lancet 2002 Jul;360(9326):2-3. [doi: 10.1016/s0140-6736(02)09358-3]
4. Nunes V, Neilson J, O'Flynn N, Calvert N, Kuntze S, Smithson H, et al. Medicines adherence: involving patients in decisions about prescribed medicines and supporting adherence. National Collaborating Centre for Primary Care. 2009. URL: https://www.nice.org.uk/guidance/cg76/evidence/full-guideline-242062957 [accessed 2020-04-15]
5. Chowdhury R, Khan H, Heydon E, Shroufi A, Fahimi S, Moore C, et al. Adherence to cardiovascular therapy: a meta-analysis of prevalence and clinical consequences. Eur Heart J 2013 Oct;34(38):2940-2948 [FREE Full text] [doi: 10.1093/eurheartj/eht295] [Medline: 23907142]
6. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-7.e81. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]

XSL•FO
RenderX

7. Wiecek E, Tonin FS, Torres-Robles A, Benrimoj SI, Fernandez-Llimos F, Garcia-Cardenas V. Temporal effectiveness of interventions to improve medication adherence: a network meta-analysis. PLoS One 2019;14(3):e0213432 [FREE Full text] [doi: 10.1371/journal.pone.0213432] [Medline: 30861014]
8. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
9. Mohammadi R, Ayatolahi Tafti M, Hoveidamanesh S, Ghanavati R, Pournik O. Reflection on mobile applications for blood pressure management: a systematic review on potential effects and initiatives. Stud Health Technol Inform 2018;247:306-310. [Medline: 29677972]
10. Pfaeffli DL, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: a systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
11. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
12. Devi R, Singh SJ, Powell J, Fulton EA, Igbinedion E, Rees K. Internet-based interventions for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2015 Dec 22(12):CD009386. [doi: 10.1002/14651858.CD009386.pub2] [Medline: 26691216]
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
14. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haqu MS, et al. Mobile applications to improve medication adherence in cardiovascular disease. National Institute for Health Research. 2019. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019121385 [accessed 2020-09-01]
15. Higgins J, Altman D, Sterne J. Chapter 8: assessing risk of bias in included studies. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0. Hoboken, New Jersey, United States: Wiley–Blackwell; 2008:1-672.
16. Review Manager (RevMan) computer program. Cochrane Training. URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman [accessed 2020-09-01]
17. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
18. Stata Statistical Software computer program, version 16. StataCorp LLC. URL: https://www.stata.com/ [accessed 2020-12-01]
19. Brath H, Morak J, Kästenbauer T, Modre-Osprian R, Strohner-Kästenbauer H, Schwarz M, et al. Mobile health (mHealth) based medication adherence measurement - a pilot trial using electronic blisters in diabetes patients. Br J Clin Pharmacol 2013 Sep 06;76 Suppl 1:47-55 [FREE Full text] [doi: 10.1111/bcp.12184] [Medline: 24007452]
20. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mhealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 06;16(7):1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
21. Fang R, Li X. Electronic messaging support service programs improve adherence to lipid-lowering therapy among outpatients with coronary artery disease: an exploratory randomised control study. J Clin Nurs 2016 Mar 01;25(5-6):664-671. [doi: 10.1111/jocn.12988] [Medline: 26522838]
22. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]
23. Goldstein CM, Gathright EC, Dolansky MA, Gunstad J, Sterns A, Redle JD, et al. Randomized controlled feasibility trial of two telemedicine medication reminder systems for older adults with heart failure. J Telemed Telecare 2014 Jun 23;20(6):293-299. [doi: 10.1177/1357633x14541039]
24. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GYH. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF app trial. Am J Med 2017 Dec;130(12):1388-1396 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
25. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
26. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]

XSL•FO
RenderX

27. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419. [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
28. Liu Y, Chen J, Lamb KV, Wu P, Chang P, Cui Y, et al. Smartphone-based self-empowerment app on secondary prevention of patients with cardiovascular disease. Stud Health Technol Inform 2019;264:1712-1713. [doi: 10.3233/shti190610]
29. Márquez Contreras E, Márquez Rivero S, Rodríguez García E, López-García-Ramos L, Carlos Pastoriza Vilas J, Baldonedo Suárez A, Compliance Group of Spanish Society of Hypertension (SEH-LELHA). Specific hypertension smartphone application to improve medication adherence in hypertension: a cluster-randomized trial. Curr Med Res Opin 2019 Jan 05;35(1):167-173. [doi: 10.1080/03007995.2018.1549026] [Medline: 30431384]
30. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: A crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446. [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
31. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809. [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
32. Ni Z, Liu C, Wu B, Yang Q, Douglas C, Shaw RJ. An mHealth intervention to improve medication adherence among patients with coronary heart disease in China: development of an intervention. Int J Nurs Sci 2018 Oct 10;5(4):322-330 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.003] [Medline: 31406843]
33. Santo K, Singleton A, Rogers K, Thiagalingam A, Chalmers J, Chow CK, et al. Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial. Heart 2019 Feb 27;105(4):323-329. [doi: 10.1136/heartjnl-2018-313479] [Medline: 30150326]
34. Sarfo FS, Treiber F, Gebregziabher M, Adamu S, Nichols M, Singh A, et al. Phone-based intervention for blood pressure control among Ghanaian stroke survivors: a pilot randomized controlled trial. Int J Stroke 2019 Aug 22;14(6):630-638. [doi: 10.1177/1747493018816423] [Medline: 30465630]
35. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;11:CD010013. [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
36. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May 15;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
37. 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 Dec;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
38. Torres-Robles A, Wiecek E, Tonin FS, Benrimoj SI, Fernandez-Llimos F, Garcia-Cardenas V. Comparison of interventions to improve long-term medication adherence across different clinical conditions: a systematic review with network meta-analysis. Front Pharmacol 2018 Dec 24;9:1454 [FREE Full text] [doi: 10.3389/fphar.2018.01454] [Medline: 30618748]
39. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_introduction.pdf?ua=1 [accessed 2020-04-15]
40. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. Clin Ther 1999 Jun;21(6):1074-1090. [doi: 10.1016/s0149-2918(99)80026-5]
41. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Dec;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
42. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
43. Desteghe L, Kluts K, Vijgen J, Koopman P, Dilling-Boer D, Schurmans J, et al. The health buddies app as a novel tool to improve adherence and knowledge in atrial fibrillation patients: a pilot study. JMIR Mhealth Uhealth 2017 Jul 19;5(7):e98. [doi: 10.2196/mhealth.7420]
44. Lee J, Evangelista LS, Moore AA, Juth V, Guo Y, Gago-Masague S, et al. Feasibility study of a mobile health intervention for older adults on oral anticoagulation therapy. Gerontol Geriatr Med 2016 Oct 07;2:- [FREE Full text] [doi: 10.1177/2333721416672970] [Medline: 28680940]
45. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014 May 01;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]
46. 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 Apr 21;17(2):52-57 [FREE Full text] [doi: 10.2196/jmir.3951]

XSL•FO
RenderX

47. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
48. Cruz Zapata B, Hernández Niñirola A, Idri A, Fernández-Alemán JL, Toval A. Mobile PHRs compliance with Android and iOS usability guidelines. J Med Syst 2014 Aug 24;38(8):1-16. [doi: 10.1007/s10916-014-0081-6] [Medline: 24957397]
49. Guo X, Yuan J, Cao X, Chen X. Understanding the acceptance of mobile health services: a service participants analysis. In: Proceedings of the 19th Annual International Conference on Management Science & Engineering. 2012 Presented at: 19th Annual International Conference on Management Science & Engineering; Sept. 20-22, 2012; Dallas, TX, USA. [doi: 10.1109/icmse.2012.6414426]
50. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010 Mar;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
51. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
52. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
53. George PP, Molina JA, Cheah J, Chan SC, Lim BP. The evolving role of the community pharmacist in chronic disease management - a literature review. Ann Acad Med Singap 2010 Nov;39(11):861-867 [FREE Full text] [Medline: 21165527]
54. Root G, Varney J. Pharmacy: a way forward for public health: opportunities for action through pharmacy for public health. Public Health England. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643520/Pharmacy_a_way_forward_for_public_health.pdf [accessed 2020-09-01]
55. George J, McNamara K, Stewart K. The roles of community pharmacists in cardiovascular disease prevention and management. Australas Med J 2011 Jun 01;4(5):266-272. [doi: 10.4066/amj.2011.698]
56. Obreli-Neto PR, Guidoni CM, de Oliveira Baldoni A, Pilger D, Cruciol-Souza JM, Gaeti-Franco WP, et al. Effect of a 36-month pharmaceutical care program on pharmacotherapy adherence in elderly diabetic and hypertensive patients. Int J Clin Pharm 2011 Aug 5;33(4):642-649. [doi: 10.1007/s11096-011-9518-x] [Medline: 21544559]
57. Jalal ZS, Smith F, Taylor D, Finlay K, Patel H, Antoniou S. Impact of pharmacy care upon adherence to cardiovascular medicines: a feasibility pilot controlled trial. Eur J Hosp Pharm 2016 Sep 02;23(5):250-256. [doi: 10.1136/ejhpharm-2015-000790] [Medline: 31156861]
58. Murray MD, Young J, Hoke S, Tu W, Weiner M, Morrow D, et al. Pharmacist intervention to improve medication adherence in heart failure: a randomized trial. Ann Intern Med 2007 May 15;146(10):714-725. [doi: 10.7326/0003-4819-146-10-200705150-00005] [Medline: 17502632]
59. DiDonato KL, Liu Y, Lindsey CC, Hartwig DM, Stoner SC. Community pharmacy patient perceptions of a pharmacy-initiated mobile technology app to improve adherence. Int J Pharm Pract 2015 Oct 09;23(5):309-319. [doi: 10.1111/ijpp.12168] [Medline: 25572628]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
DBP: diastolic blood pressure
HCP: health care professional
LDL-C: low-density lipoprotein cholesterol
mHealth: mobile health
PAM: patient activation measure
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SBP: systolic blood pressure
TC: total cholesterol

XSL•FO
RenderX

Edited by G Eysenbach; submitted 08.09.20; peer-reviewed by T Miron-Shatz, E Wiecek, P Cook; comments to author 28.11.20; revised version received 08.03.21; accepted 11.04.21; published 25.05.21.

Please cite as:
Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, Jalal Z
Mobile Apps to Improve Medication Adherence in Cardiovascular Disease: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(5):e24190
URL: https://www.jmir.org/2021/5/e24190
doi:10.2196/24190
PMID:34032583

©Shahd Al-Arkee, Julie Mason, Deirdre A Lane, Larissa Fabritz, Winnie Chua, M Sayeed Haque, Zahraa Jalal. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Digital Health Strategies for Cervical Cancer Control in Low- and Middle-Income Countries: Systematic Review of Current Implementations and Gaps in Research

Andrea H Rossman[1], BSE; Hadley W Reid[2], BA; Michelle M Pieters[3], MSc; Cecelia Mizelle[3], BSc; Megan von Isenburg[4], MSLS; Nimmi Ramanujam[1,3], PhD; Megan J Huchko[3,5], MPH, MD; Lavanya Vasudevan[3,6], MPH, PhD, CPH

[1]Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC, United States

[2]Duke School of Medicine, Durham, NC, United States

[3]Duke Global Health Institute, Durham, NC, United States

[4]Medical Center Library, Duke University, Durham, NC, United States

[5]Department of Obstetrics and Gynecology, Duke University School of Medicine, Durham, NC, United States

[6]Department of Family Medicine and Community Health, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Lavanya Vasudevan, MPH, PhD, CPH
Department of Family Medicine and Community Health
Duke University School of Medicine
2200 W. Main Street
Suite 600
Durham, NC, 27705
United States
Phone: 1 9196131423
Email: lavanya.vasudevan@duke.edu

Abstract

Background: Nearly 90% of deaths due to cervical cancer occur in low- and middle-income countries (LMICs). In recent years, many digital health strategies have been implemented in LMICs to ameliorate patient-, provider-, and health system–level challenges in cervical cancer control. However, there are limited efforts to systematically review the effectiveness and current landscape of digital health strategies for cervical cancer control in LMICs.

Objective: We aim to conduct a systematic review of digital health strategies for cervical cancer control in LMICs to assess their effectiveness, describe the range of strategies used, and summarize challenges in their implementation.

Methods: A systematic search was conducted to identify publications describing digital health strategies for cervical cancer control in LMICs from 5 academic databases and Google Scholar. The review excluded digital strategies associated with improving vaccination coverage against human papillomavirus. Titles and abstracts were screened, and full texts were reviewed for eligibility. A structured data extraction template was used to summarize the information from the included studies. The risk of bias and data reporting guidelines for mobile health were assessed for each study. A meta-analysis of effectiveness was planned along with a narrative review of digital health strategies, implementation challenges, and opportunities for future research.

Results: In the 27 included studies, interventions for cervical cancer control focused on secondary prevention (ie, screening and treatment of precancerous lesions) and digital health strategies to facilitate patient education, digital cervicography, health worker training, and data quality. Most of the included studies were conducted in sub-Saharan Africa, with fewer studies in other LMIC settings in Asia or South America. A low risk of bias was found in 2 studies, and a moderate risk of bias was found in 4 studies, while the remaining 21 studies had a high risk of bias. A meta-analysis of effectiveness was not conducted because of insufficient studies with robust study designs and matched outcomes or interventions.

Conclusions: Current evidence on the effectiveness of digital health strategies for cervical cancer control is limited and, in most cases, is associated with a high risk of bias. Further studies are recommended to expand the investigation of digital health strategies for cervical cancer using robust study designs, explore other LMIC settings with a high burden of cervical cancer (eg, South America), and test a greater diversity of digital strategies.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e23350) doi:10.2196/23350

KEYWORDS

cervical cancer; digital health; mobile phones; low- and middle-income countries; colposcopy; uterine cervical neoplasms; telemedicine or mobile apps; cell phones; developing countries

Introduction

Background

Annually, 311,000 women die from cervical cancer worldwide, with 90% of the deaths occurring in low- and middle-income countries (LMICs). [1]. To reduce this high burden of mortality, it is critical to implement and scale sustainable and effective cervical cancer control programs in LMICs. However, cervical cancer control programs in LMICs must overcome individual-, provider-, and health system–level bottlenecks to health service delivery, access, and utilization [2-5]. In the last decade, health systems in LMICs have taken advantage of the increasing prevalence of digital technologies, particularly mobile phones, to circumvent some of the bottlenecks in cervical cancer control. However, systematic reviews of the effectiveness of such digital health strategies in LMIC settings are lacking, especially regarding how such strategies may improve the delivery, access, and utilization of cervical cancer control programs [4].

Primary prevention strategies for cervical cancer control focus on vaccination against human papillomavirus (HPV), whereas secondary prevention strategies focus on early screening and treatment of precancerous lesions [6,7]. Where implementation of preventive services is inadequate, cervical cancers are typically detected at advanced stages, further contributing to high mortality rates [8,9]. Commonly used screening methods include visual inspection with acetic acid (VIA), visual inspection with Lugol iodine, and HPV DNA testing [10,11]. Although an initial cervical cancer screening is possible at primary health facilities, or even at the community level by frontline health workers, follow-up procedures, diagnosis, and treatment require access to trained medical providers (eg, expert colposcopists). The treatment of invasive cancer also requires secondary or tertiary health facilities with specialized equipment for surgery, chemotherapy, and radiation [11]. The management of invasive cancer includes long-term retention of women in care, with regular follow-ups and palliative care as applicable. To be effective, cervical cancer control programs must not only achieve high rates of screening coverage among eligible women but also ensure that women who screen positive receive timely treatment and support for long-term management [12]. There is a lack of systematic reviews to understand how digital health strategies affect cervical cancer control in LMICs across the continuum spanning from prevention to palliative care.

There is growing interest from local, national, and global stakeholders in integrating digital health strategies with cervical cancer control programs in LMICs. For instance, the Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020 highlights the use of digital health strategies for health education, promotion, and communication, especially in populations with low literacy and health awareness [13]. Examples of using digital health for cervical cancer control include the World Health Organization and the International Telecommunications Union's *Be he@lthy, Be mobile* initiative, which launched a mobile phone text messaging campaign to improve the awareness of cervical cancer screening in Zambia [14]. In another instance, the Ministry of Health in Peru successfully completed a national pilot program of using text messages to notify women about their HPV screening results [15]. In light of the expanding local, national, and global efforts, evidence of successful implementation and impact is needed to drive further research on and development of digital strategies for cervical cancer control in LMICs.

Currently, much of the published literature describing the use of digital strategies for cervical cancer control comes from high-income settings [16-19]. However, a previous systematic review examined the effectiveness of digital strategies for cervical cancer control in LMICs [20]. In this review, authors identified 8 eligible studies, most of which lacked rigorous study designs and evidence. With the increasing use of digital health strategies in LMICs, an updated review is necessary to evaluate the contextual effectiveness of such strategies for cervical cancer control as compared with high-income settings. Furthermore, the synthesis of key implementation challenges and opportunities for future research is needed to prioritize and increase the success of digital health strategies for cervical cancer control in LMICs.

Objectives

The primary objectives of the review were to assess the effects of using mobile devices on the following:

- Facilitate task shifting for cervical cancer screening, treatment of precancerous lesions, and management of LMICs as compared with usual care.
- Reduce delays in postscreening treatment initiation among women in LMICs as compared with usual care.
- Assess and improve cervical cancer knowledge or awareness among women in LMICs as compared with usual care.

The secondary objectives were to describe the following:

- Digital health strategies used for cervical cancer control, including those used to visualize the cervix using a mobile device at the point of care in LMICs.
- Challenges associated with the implementation of digital health strategies for cervical cancer control in LMICs.

Methods

Protocol and Registration

The review protocol was registered with the PROSPERO database for prospectively registered systematic reviews (protocol #CRD42017071560). Deviations from the registered protocol and unused methods are included in Multimedia Appendix 1. The findings of this systematic review are reported

XSL•FO
RenderX

in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and are specified in Multimedia Appendix 2 [21].

Eligibility Criteria

Participants

For the assessment of the primary and secondary review objectives, we included studies of all cadres of health care workers (eg, medical doctors, nurses, midwives, community health workers) providing any cervical cancer screening, treatment, and management services, as well as women (of all ages) receiving any cervical cancer education, screening, treatment, and management services.

Interventions

For the assessment of primary and secondary review objectives, we included any intervention wherein mobile devices were used in the screening, treatment, and management of cervical cancer. These included studies in which mobile devices were used to do the following:

- Facilitate the visualization of the cervix (cervicography).
- Facilitate communication between health care providers for diagnosis, consultation, or referral.
- Provide training or support to health care providers, especially in the context of task shifting.
- Communicate with patients to provide appointment reminders, test results, disease progression monitoring, coordinate services, etc.
- Improve cervical cancer knowledge or awareness among women.
- Facilitate cervical cancer screening, treatment of preinvasive lesions or cervical cancer management.

We excluded the following types of studies:

- Where the intervention used a nonmobile phone device (eg, digital camera, television) to visualize the cervix.
- Where the intervention exclusively focused on the use of mobile devices for providing education about HPV vaccination or to improve its coverage, because these activities generally fall within the purview of immunization programs. In addition, other published systematic reviews have examined the efficacy of mobile phone interventions for HPV vaccination [22-24].
- Where the intervention was only described or conceptualized without actual implementation or evaluation.
- Where abstracts were only published in languages other than English and for which published English translations were not available.
- Where full texts were not available.
- Where the publications were reviews, systematic reviews, conference proceedings, blogs, reports, or other nonpeer-reviewed sources.

Comparators

Details of the comparator group were extracted from the studies and included in the analysis of primary and secondary review objectives. The comparison was usual care, which could include the use of traditional colposcopy and nondigital, paper-based strategies for data collection, communication, and dissemination in cervical cancer control programs.

Outcomes

For the review of the primary review objectives, the outcomes of interest were as follows:

1. Coverage and timeliness of cervical cancer screening, treatment, and management.
2. Cervical cancer knowledge or competency (ie, training) of health worker.
3. Cervical cancer knowledge or awareness among women.

Outcome data were not reviewed quantitatively because of the descriptive nature of the secondary review objectives.

Study Design

We included randomized and nonrandomized controlled study designs (controlled before-and-after studies with at least 2 intervention sites or interrupted time-series studies) in the qualitative synthesis and meta-analysis of the primary review objectives. We included all study designs for a narrative review of the secondary review objectives.

Settings

For both the primary and secondary review objectives, we included studies from any country listed as low or middle income according to the World Bank Group classification [25].

Information Sources

Electronic Searches

The following 5 electronic databases were searched for studies published in English: PubMed, Embase, Web of Science, Scopus, and CINAHL. We included studies from January 1, 1992, to September 19, 2020 (date of search), with 1992 being the year the first commercial text message was sent.

Search Strategy

A systematic search strategy (Multimedia Appendix 3) was developed, including a detailed search string comprising terms from 3 broad categories, namely, digital health, cervical cancer, and LMICs. The search terms were customized for each of the 5 electronic databases. A Google search was conducted with an abbreviated search string, as described in Multimedia Appendix 3, and the first 100 search results were analyzed.

Study Selection

Search records were imported into the reference management software and duplicates were removed. The titles and abstracts of the records were screened according to the predefined eligibility criteria. For each record, 2 reviewers discussed and resolved any ambiguities in screening outcomes during abstract and full-text screening. Interrater reliability was assessed by 2 reviewers independently screening a 10% sample of sources for inclusion based on abstract evaluation. A Cohen κ value of 0.6 was predetermined as an acceptable interrater agreement.

Data Collection Process and Items

We used a structured template, which was adapted based on the template from Cochrane, to extract relevant data from each

XSL•FO
RenderX

included study [26]. Primary data items extracted using the template included study location, sample size, population, study duration and design, cervical cancer service or procedure, the description of digital health intervention, conclusions, study limitations, and information for assessing the risk of bias. In addition, quantitative outcome data were extracted from the studies included in the review of the primary objectives.

Risk of Bias in Individual Studies

We assessed the risk of bias for all studies using the Effective Public Health Practice Project's Quality Assessment Tool for Quantitative Studies [27].

Additional Analyses

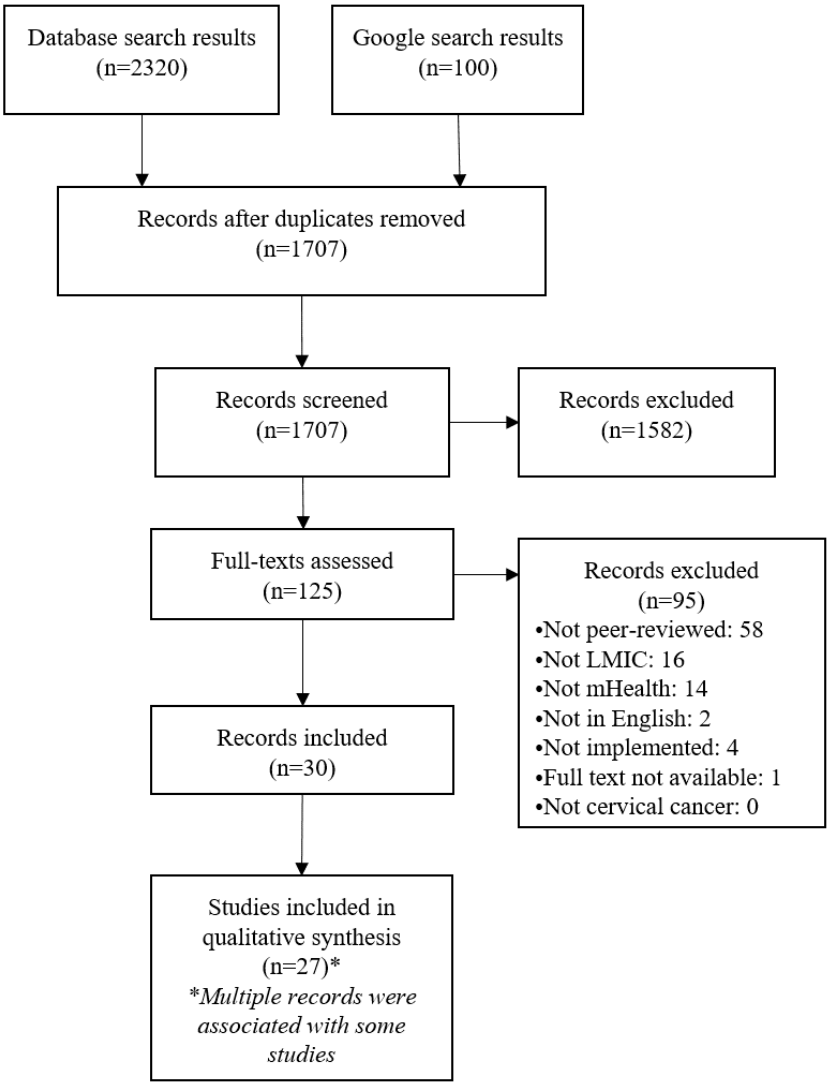
We assessed the quality of data reporting for each included study using the mobile health (mHealth) evidence reporting and assessment (mERA) checklist [28]. This 16-item checklist aims to enhance the replicability of mHealth interventions by promoting the complete reporting of content, context of implementation, and technical features in peer-reviewed publications.

Results

Study Selection

Figure 1 shows the results of the study selection process. The systematic search yielded 1707 nonduplicate studies. The overall interrater agreement for screening a 10% sample was 98.8%, which corresponded to a Cohen κ value of 0.76 [29]. Following the abstract screening, 125 records were evaluated using their full text and 95 of those were excluded. Records were excluded after full-text review mainly because they did not describe a peer-reviewed study or because the intervention was conducted in a non-LMIC setting. In total, 30 records corresponding to 27 unique studies [30-59] were included in the final review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of systematic search results. LMIC: low- and middle-income country; mHealth: mobile health.



Study Characteristics

Characteristics of included studies are summarized in Multimedia Appendix 4 [30-59].

Study Design

In total, 4 included studies were randomized controlled trials (RCTs) [38,40,47-49,59]. These studies met the criteria for inclusion in the review of the primary objectives, that is, the

XSL•FO
RenderX

effect of digital health strategies on cervical cancer control. A fifth study by Lima et al [41] used random allocation but was excluded from the review of the primary objectives because of the lack of a control group. Huchko et al [39] reported findings of a mixed methods analysis from a cluster RCT; however, the assignment of the digital health strategy was not randomized; thus, it did not meet the criteria for inclusion in the review of primary objectives. All 27 studies were included in the review of the secondary objectives, that is, to describe the digital health strategies used for cervical cancer control in LMICs and the challenges associated with their implementation. Nonrandomized studies used quasi-experimental [32-34,41], cross-sectional [30,31,35-37,46,50,51,53,56,57], or other [52,54,55] study designs. In total, 4 studies [42-45] described program implementation.

Participants

The studies included 23,393 women and 152 health care workers. Types of health care workers included clinicians (eg, gynecologists, colposcopists, and assistant medical officers), facility-based nurses, and community health nurses.

Intervention

Many interventions targeted health workers and implemented the use of mobile phones for digital cervicography (ie, imaging of the cervix) or for digital patient data collection [30-32,35-37,42-46,50,51,53-57]. Interventions targeting women were focused on building knowledge or awareness about cervical cancer, delivering reminders to promote the uptake of cervical cancer services or for the notification of test results [33,34,38-41,47-49,52,58,59].

Comparators

Most of the included studies lacked a parallel control group. Apart from the 4 RCTs and the cluster randomized trial, the only other studies with a control group were Caster et al and Romli et al [34,58].

Outcomes

Studies using digital cervicography primarily reported diagnostic reliability outcomes such as interrater agreement and sensitivity or specificity as compared with a gold standard diagnostic test such as histology [30,31,35,46,50,51,53-57]. Among the 4 RCTs included in the review of primary objectives, 1 study [40] measured improvement in patient knowledge and all 4 [38,40,47-49,59] measured the uptake of cervical cancer services postintervention. One RCT study [49] included a cost-effectiveness analysis. For the secondary review objectives, all 27 studies included a description of the digital health strategy used and implementation challenges.

Settings

The review focused on LMICs: 19 studies were conducted in sub-Saharan Africa [30,31,34-39,42-46,50,52,54-56,59], 7 in Asia [32,33,40,47-49,51,53,58], and 1 in South America [41].

Risk of Bias Within Studies

The analysis results of the risk of bias within the studies are presented in Table 1 (Effective Public Health Practice Project Quality Assessment tool for quantitative studies). Most of the studies (21/27, 78%) were assessed to have a high risk of bias. A total of 2 studies (Erwin et al [38] and Romli et al [58]) were assessed to have a low risk of bias, whereas 4 other studies [34,40,53,59] were assessed to have a moderate risk of bias.

XSL•FO
RenderX

Table 1. Risk of bias assessment for included studies using the Effective Public Health Practice Project's quality assessment tool for quantitative studies[a].

Study	Selection bias	Study design	Confounders	Blinding	Data collection method	Withdrawals and dropouts	Global rating
Asgary et al (2016) [30]; Asgary et al (2019) [31]	1	3	3	2	1	1	3
Bhatt et al (2018) [32]	3	3	3	3	3	3	3
Caster et al (2015) [34]	1	2	3	2	2	1	2
Catarino et al (2015) [35]	2	3	3	2	1	N/A[b,c]	3
Devi et al (2018) [33]	3	3	3	3	2	3	3
Erwin et al (2019) [38]	1	1	2	2	2	1	1
Gallay et al (2017) [36]	2	3	1	2	3	1	3
Huchko et al (2019) [39]	1	3	2	3	2	2	3
Khademolhosseini et al (2017) [40]	2	1	1	3	1	1	2
Lima et al (2017) [41]	2	1	1	1	3	3	3
Linde et al (2020) [59]	1	1	1	2	3	1	2
Littman-Quinn et al (2013) [42]	2	3	3	3	3	3	3
Ndlovu et al (2014) [43]	2	3	3	3	1	3	3
Parham et al (2010) [44]	2	3	3	3	3	3	3
Peterson et al (2016) [45]	1	3	1	3	3	N/A	3
Quercia et al (2018) [37]	2	3	N/A	3	1	1	3
Quinley et al (2011) [46]	1	3	3	2	1	N/A	3
Rashid et al (2013) [47]; Rashid and Dahlui (2013) [48]; Rashid et al (2014) [49]	1	1	3	2	3	N/A	3
Ricard-Gauthier et al (2015) [50]	2	3	3	2	1	2	3
Romli et al (2020) [58]	1	1	1	2	1	1	1
Sharma et al (2018) [51]	2	3	N/A	3	2	2	3
Swanson et al (2018) [52]	1	2	1	3	3	3	3
Taghavi et al (2018) [53]	1	3	N/A	2	1	1	2
Tran et al (2018) [54]	2	3	3	2	3	3	3
Urner et al (2017) [55]	2	3	1	2	2	3	3
Yeates et al (2016) [56]	2	3	3	3	3	3	3
Yeates et al (2020) [57]	2	3	3	2	3	2	3

[a]Scores of 1, 2, and 3 indicate low, moderate, and high risks of bias, respectively. The risk of bias was assessed cumulatively for studies with multiple sources, for example, Asgary et al [30,31].

[b]N/A: not applicable.

[c]Criteria were not applicable based on a skip pattern in the Effective Public Health Practice Project tool.

Results of Individual Studies

The study outcomes included in the review of the primary objectives are presented in Table 2.

XSL•FO
RenderX

Table 2. Outcomes of randomized controlled trial studies included in the review of primary objectives.

Study and participants	Outcome	Result	Summary of risk of bias[a]
Erwin et al (2019) [38], N=851 women			
• 281 controls • 272 SMS messaging • 298 SMS messaging+e-voucher	Cervical cancer screening attendance within 60 days of randomization (combined for women from rural and urban settings)	• Women in the SMS messaging group had 3.0 higher adjusted odds of attendance as compared with women in the control group • Women in the SMS messaging+e-voucher group had 4.7 higher adjusted odds of attendance as compared with women in the control group • Women in the SMS messaging+e-voucher group had 1.5 times higher adjusted odds of attendance compared with women in the SMS messaging group	The overall risk of bias was assessed to be low
Khademolhosseini et al (2017) [40], N=95 women			
• 47 control • 48 intervention	Mean difference in pre- and posttraining knowledge among women in the intervention group as compared with those in the control group measured immediately and 3 months after SMS messaging–based training	• Women in the intervention group had a mean increase in the knowledge of 8.18 points from baseline as compared with a mean increase of 0.27 points from baseline in the control group immediately posttraining • Women in the intervention group had a mean increase in the knowledge of 8.35 points from baseline as compared with a mean increase of 0.17 points in the control group at 3 months of posttraining	The overall risk of bias was assessed to be moderate
• 47 control • 48 intervention	Uptake of Pap test within 3 months of training in the intervention group compared with control group	• At 3 months of posttraining, only 4 (5.8%) participants of the control group as compared with 23 (47.9%) participants of the intervention group had received a Pap test	The overall risk of bias was assessed to be moderate
Linde et al (2020) [59], N=705 women			
• 347 standard of care (control) • 358 standard of care+text message	The attendance rate of follow-up cervical cancer screening among HPV[b]-positive women	• Compared with standard of care, a written appointment card, the addition of one-way text messages had no effect on follow-up cervical cancer screening among HPV-positive women	The overall risk of bias was assessed to be moderate
Rashid and Dahlui (2013) [48], N=1000 women			
• 250 letters • 250 registered letters • 250 SMS messaging • 250 phone calls	The uptake of Pap test in response to recall through phone call as compared with recall by letter	• Compared with women receiving recall by letter, those receiving recall by phone call had 2.38 times higher odds of receiving a Pap smear	The overall risk of bias was assessed to be high
• 250 letters • 250 registered letters • 250 SMS messaging • 250 phone calls	The uptake of Pap test in response to recall through SMS messaging as compared with recall by letter	• Compared with women receiving recall by letter, those receiving recall by SMS messaging had no significant change in the odds of receiving a Pap smear	The overall risk of bias was assessed to be high

[a]Assessed using the Effective Public Health Practice Project risk of bias assessment tool for quantitative studies.

[b]HPV: human papillomavirus.

Synthesis of Results

Findings From the Review of Primary Objectives: Effectiveness of Digital Health Strategies for Cervical Cancer Control

We did not conduct a meta-analysis of the included studies because of variations in the nature of the interventions, outcome measures, and small number of studies included in the primary review of the effect of mobile devices. The only included study with a low risk of bias found that SMS behavior change communication messaging in conjunction with transportation e-voucher led to an increased uptake of cervical cancer screening (Table 2) [38].

XSL•FO
RenderX

Findings From the Review of Secondary Objectives: Description of Strategies Used

We adapted a published mHealth framework for noncommunicable diseases to facilitate the narrative synthesis of the digital health strategies used in the included studies [4,60]. Framework adaptation allowed for the mapping of included studies to various stages of the cervical cancer control cascade (primary prevention, secondary prevention, treatment, and palliation) as well as the key individual, provider, and health system challenges (knowledge, access, quality, and continuity of care) addressed by the included studies. The mapping results are presented in Table 3. Only the digital health components of the study interventions were mapped and any nondigital components (community sensitization, paper educational booklets administered before digital reminders, etc) were excluded. Many studies addressed multiple challenges; hence, digital health strategies were mapped to all the applicable challenges in the framework.

Table 3. Landscape of digital health strategies for cervical cancer prevention and control.

Individual-, provider- and health system–level challenges in cervical cancer control	Stages in cervical cancer prevention and control				
	Primary prevention	Secondary prevention		Treatment and palliative care	
	HPV[a] vaccination	Screening (study)	Treatment of precancerous lesions (study)	Treatment	Palliative care
Knowledge and awareness					
Low knowledge of HPV or cervical cancer	Not within the scope of this review	5 [34,38,40,58,59]	1 [34]	0	0
Low knowledge of cervical cancer screening or treatment services	Not within the scope of this review	6 [34,38,40,41,58,59][b]	1 [34]	0	0
Access to care					
Low access to health facilities or cervical cancer services	Not within the scope of this review	1 [38]	0	0	0
Low availability of appropriate and accurate screening or treatment methods	Not within the scope of this review	15 [30,31,35,36,42-46,50,51,53-57][c]	0	0	N/A
Low access to experts	Not within the scope of this review	15 [30,31,35,36,42-46,50,51,53-57]	0	0	0
Financial barriers	Not within the scope of this review	1 [38]	0	0	0
Continuity of care					
A low uptake of follow-up services	Not within the scope of this review	7 [33,39,41,47-49,52,58,59]	0	0	N/A
Quality of care					
A lack of training opportunities for health workers	Not within the scope of this review	10 [30,31,35,36,42-44,46,51,56,57][c]	0	0	0
Poor data availability	Not within the scope of this review	19 [30-32,35-37,39,42-46,50-57]	2 [32,45]	0	0

[a]HPV: human papillomavirus.

[b]Represents the number of included studies mapped to each category. For example, we found 3 studies that aimed to increase the demand for screening by increasing women's knowledge of human papillomavirus or cervical cancer.

[c]Some studies were associated with multiple records, for example, Asgary et al [30,31].

A majority of the included studies used digital health strategies for secondary prevention; only Caster et al [34] focused on educating women about treatment, whereas Bhatt et al [32] and Peterson et al [45] collected data related to postscreening treatment. None of the included studies focused on cervical cancer treatment or palliative care. Studies related to primary prevention (HPV vaccination) were not within the scope of the study. Although most included studies focused on increasing screening or treatment among all eligible women, some studies focused on high-risk populations, and in particular, HPV-positive women [35-37,50,53,54,59].

1. *Knowledge and awareness*: in total, 6 studies mapped to this domain and included digital health strategies to educate women about HPV, cervical cancer, and cervical cancer prevention or treatment services. These studies

XSL•FO
RenderX

[34,38,40,41,58,59] focused primarily on increasing the screening uptake: Khademolhosseini et al [40] used an educational intervention delivered via an instant messaging platform called Telegram. They used a diverse range of content including text messages, posters, infographics, podcasts, and video tutorials. Lima et al [41] tested a telephone intervention focused on increasing the patient's knowledge of cervical cancer and Pap smears. In the study by Erwin et al [38], women in one study arm received 15 behavior change communication messages via SMS messages that were designed to increase their knowledge and awareness about cervical cancer screening. Caster et al [34] implemented a tablet-based cervical cancer education program to facilitate patient education about screening. Opportunities for interaction with the educational content were through quizzes, or while navigating the content on the tablet. Caster et al [34] also included educational content related to treatment and was the only study to do so.

2. *Access to care*: only one study, by Erwin et al [38], used digital health to reduce transportation and financial barriers to health care access. Women randomized to one of the study arms received e-vouchers through their mobile phone, which covered the costs of a return trip to the health facility. In total, 15 studies used smartphones for facilitating digital cervicography and visualizing the cervix during VIA at the point of care [30,31,35,36,42-46,50,51,53-57]. In total, 7 studies included visual inspection with Lugol iodine images in addition to VIA images [35-37,50,53-55]. Although most studies used the mobile phone camera for image acquisition, Peterson et al [45] and Taghavi et al [53] used smartphone attachments for enhanced cervix visualization. These studies allowed for task shifting, as mobile devices were used to acquire cervical images, record diagnosis, and receive or compare diagnoses with remote experts. Images and relevant patient data were shared with the experts via a text/multimedia message service or by uploading data to a web-based database. Parham et al [44] used an automated text messaging system to reduce the time between screening, diagnosis, and treatment by sending a text message requesting expert review of an image while the patient was still in the clinic. Yeates et al [56,57] used WhatsApp to send patient images, nurse diagnosis, and treatment plans for expert review and subsequently developed the SEVIA app to accomplish these tasks.
3. *Continuity of care*: in total, 7 studies tested digital strategies for reminding or recalling patients for follow-up services [33,39,41,47-49,52,58,59]. These studies focused on women who had already received some cervical cancer services. All 7 studies tested SMS-based reminders or recall messages. In total, 3 studies [39,47-49,52] also tested phone calls. Other nondigital modalities included letters, registered letters, home visits, and return visits to the clinic.
4. *Quality of care*: studies mapped to this domain focused on 2 key applications: training health workers and improving data availability. In most cases, the training covered the acquisition of good-quality images using cervicography and expert feedback to improve the accuracy of diagnosis [30,31,35,36,42-44,46,51,56,57]. In Botswana, Littman-Quinn et al [42] provided additional ongoing medical education content to health workers via tablet devices to complement in-service training. Studies using digital cervicography were also mapped to this domain as the documentation of cervical images during VIA can improve data availability for patient management, and hence, improve the quality of care. Two studies, Bhatt et al [32] and Peterson et al [45], collected treatment information in addition to information about screening results. In the study by Bhatt et al [32], a SIM-based app was loaded on feature phones provided to trained nurses. The nurses used a menu-based protocol on the feature phone for entering data.

Findings From the Review of Secondary Objectives: Implementation Challenges

In the included studies, the authors described several technical challenges in implementing digital health strategies for cervical cancer control in LMICs. These challenges were based on *lessons learned* and, in a few cases [31,32], on formal qualitative data collected during the study. In addition to the technical challenges, many studies described inadequacies in the underlying health system resources (eg, availability of pelvic exam rooms and lack of supplies for cryotherapy in community health centers [30]), which impacted study implementation. Digital health strategies were not implemented in a vacuum; several studies described the need for increasing the community knowledge and awareness of cervical cancer as well as reducing stigma and fear related to cancer screening/diagnosis in parallel with the implementation of digital health strategies [30,32,37,38,41,44,45,51]. One study also described negative attitudes toward cervical cancer screening among health care workers as an issue [51]. Some authors emphasized the need for strong partnerships and stakeholder support for the success and future sustainability of digital health strategies [42,43,45,51]. Table 4 shows the technical challenges in implementing digital health strategies for cervical cancer control.

XSL•FO
RenderX

Table 4. Description of implementation challenges for digital health strategies for cervical cancer control.

Implementation challenge	Included studies	Description and examples

High training requirements	[30,31,34,35,43,44,51,54]	• Pretraining lasting several weeks, providing supplemental training for augmenting skills, refresher training to minimize loss of skills, and the availability of experts to provide ongoing feedback for using digital cervicography
		  • Catarino et al [35]: 5 weeks of training on digital cervicography to medical students
		  • Asgary et al [30,31]: additional one-on-one training to community health nurses in cases where the digital image quality was low
		  • Ndlovu et al [43]: high training requirements on the touchscreen features
		  • Caster et al [34]: Users had limited previous experience with technology (but needed very little support for using the tablet device).

Technology-specific challenges	[30,32,36,42-44,46,50,54,55]	• Procurement of appropriate technology
		  • Challenges with the availability of technology options in the study area in preparation for and during the study [30,32]
		  • Considerations related to finding mobile phone cameras with high image quality and zoom capabilities [30,43,46,54,55]
		  • The use of high-pixel smartphone cameras were associated with better reported quality of images [50]; however, Tran et al [54] suggested that the quality was inferior to colposcopy images.
		  • Parham et al [44]: need to send cameras out of the country for repairs was a challenge
		• Software and hardware issues: software "bugs," crashing of apps, device malfunctions, and an insufficient battery life [36,42-44]
		  • Gallay et al [36]: loss of patient data due to the unexpected crash of their data collection app
		• Data security issues
		  • Littman-Quinn et al [42]: security breach (attack by an anonymous hacker) lead researchers to increase data security to the level of compliance described in the Health Insurance Portability and Accountability Act
		• Software updates
		  • Bhatt et al [32]: challenges in updating a SIM-based app requiring project staff to collect all SIM cards to update the app
		  • Littman-Quinn et al [42]: challenges with communication when the technical team was based in a different country and did not speak the same language as the end users, reflecting the need for local technology development capacity for the sustainability of digital health strategies

Infrastructure challenges	[31,32,42,43,45,46]	• Issues with network coverage and electrical outages as limitations to widespread implementation
		  • Bhatt et al [32]: challenges faced by nurses from hillier communities in sending patient data and receiving acknowledgment of report submission
		  • Yeates et al [56]: health care workers given solar-powered chargers and light sources for anticipated power outages and to allow use of digital cervicography in off-site settings

Challenges with technology reach	[39,47-49]	• Rashid et al [47-49]: letters were more likely to be successfully and reliably delivered to patients than phone text messages or calls because of connectivity and coverage issues (incorrect phone numbers and nonresponse) • Rashid et al [47-49] and Huchko et al [39]: direct communication through phone call encouraged more women to seek screening (both studies had a high risk of bias, limiting our confidence in their findings)
Inequitable access to technology	[38,39,52]	• Exclusion of women without mobile phones from digital health intervention components in some studies

Additional Analysis: Quality of Digital Health Reporting

The quality of reporting scores according to the mERA checklist is summarized in Multimedia Appendix 5 [30-59]. The mERA scores ranged from 1 to 10, out of a maximum possible score of 16. The mERA checklist items frequently described in the studies included the technology platform and details of intervention delivery. Poorly described items were related to digital health infrastructure necessary for implementation, interoperability with other existing digital health systems, usability testing during development, and intervention

XSL•FO
RenderX

replicability, with 6 or fewer included studies describing these characteristics.

Discussion

Summary of Evidence

Our review findings show that the majority of the studies in LMICs used digital health strategies to facilitate the screening and treatment of precancerous lesions (ie, secondary prevention) as compared with the treatment of invasive cancer or palliative care. Even though our search attempted to include any relevant studies since 1992, the date of the earliest included study was 2010. Within the realm of secondary prevention, strategies focused on improving women's knowledge and awareness about cervical cancer, increasing access to cervical cancer services, improving the training of health workers and availability of data, and ensuring the continuity of care. Key challenges in implementing digital health strategies for cervical cancer control were related to the high burden of training, technology-specific issues, infrastructure challenges, challenges with technology reach, and inequitable access to technology among target users. We were unable to determine the quantitative effect of digital health strategies on cervical cancer control because of the small numbers and inadequate quality of studies for meta-analysis. Only one randomized controlled study identified in this review with a low risk of bias found that SMS behavior change communication messaging in conjunction with a transportation e-voucher leads to an increased uptake of cervical cancer screening. Most of the studies included in the review had a high risk of bias and were rated poorly in terms of the quality of reporting of the digital health strategy.

Implications for Research and Practice

This review identified several gaps in the literature. These gaps are summarized below, along with their implications for research and practice:

1. Improve the evidence base for the effectiveness of digital health strategies for cervical cancer control: there is insufficient evidence related to the effectiveness of digital health strategies for cervical cancer screening and treatment. The included studies implemented digital health strategies mostly for secondary prevention, and there are opportunities to investigate the use of digital health for cervical cancer treatment and palliative care. Other bottlenecks in the cervical cancer control cascade that may benefit from using digital health include improving access to health facilities (eg, through the use of digital telemedicine), reducing financial barriers (eg, provision of phone vouchers), and supporting disease management among women diagnosed with invasive cancer (eg, using digital knowledge interventions).
2. Use more rigorous study designs: among the 27 included studies, only 4 studies used an RCT design [38,40,47-49,59] and only one of these studies had a low risk of bias [38]. The high or moderate risk of bias among the remaining studies limited our confidence in their findings. Future studies should consider using rigorous study designs that minimize the risk of bias.
3. Improve the reporting of digital health strategies in the literature: in our review, only 5 studies [35-37,39,54,57,59] met the mERA checklist item on replicability, indicating that many of the digital health strategies identified in this review would be difficult to replicate and re-evaluate based on published information. The use of reporting checklists ensures that all relevant information is presented to readers to assist with study reproducibility.
4. Expand research on LMIC settings in Asia and South America: a majority (70%) of the included studies took place in sub-Saharan Africa, indicating an opportunity to expand research to other LMIC settings with a high burden of cervical cancer.

Limitations

Our synthesis of the literature is limited by the availability of peer-reviewed reports of digital health strategies for cervical cancer control. We tried to mitigate this limitation by using a systematic search strategy and searching 5 large databases and Google to identify relevant studies. We did not search any trial databases for ongoing studies or to examine other gray literature sources. Hence, we may have missed ongoing implementations and emerging data on using digital health for cervical cancer control.

Conclusions

There is insufficient evidence to determine the effectiveness of digital health strategies for cervical cancer control in LMICs. The only RCT study identified in this review with a low risk of bias found that SMS behavior change communication messaging in conjunction with a transportation e-voucher led to an increased uptake of cervical cancer screening. Future efforts are needed to investigate the use of digital health strategies across the cervical cancer control continuum and in LMIC settings outside of sub-Saharan Africa.

Acknowledgments

The authors would like to thank Patricia G Moorman, PhD, for providing feedback on the early drafts of the manuscript and Christina Makarushka, MPH, for coordinating this review. This study was funded by a pilot grant from the Duke Cancer Initiative and 1R01CA195500 Quick Trials: a viable solution for a see and treat paradigm for cervical precancer in Africa, and 1R01CA19338001 AIP: Culturally Appropriate Screening and Diagnosis of Cervical Cancer in East Africa to NR. HWR and LV receive funding from the National Center for Advancing Translational Sciences of the National Institutes of Health under award numbers TL1TR002555 and KL2TR002554, respectively. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

XSL•FO
RenderX

Authors' Contributions

LV conceived the study and protocol. AHR, LV, and MVI developed the search terms. The medical librarian (MVI) conducted the systematic search. AHR, LV, CM, and HWR screened studies for inclusion, and AHR, LV, MMP, CM, and HWR extracted data and completed the bias analysis. AHR developed figures and tables. AHR and LV cowrote the first draft of the manuscript. All authors have revised the manuscript and approved the final version for submission.

Conflicts of Interest

One review author (MJH) was listed as an author on 2 included studies and did not participate in the risk of bias assessment of any included study. The authors declare no other conflicts of interest.

Multimedia Appendix 1
Protocol deviations and unused methods.
[DOCX File , 14 KB - jmir_v23i5e23350_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 19 KB - jmir_v23i5e23350_app2.docx]

Multimedia Appendix 3
Systematic search strategy, customized by database.
[DOCX File , 26 KB - jmir_v23i5e23350_app3.docx]

Multimedia Appendix 4
Characteristics of included studies.
[DOCX File , 28 KB - jmir_v23i5e23350_app4.docx]

Multimedia Appendix 5
Quality of reporting of including studies using the mHealth evidence reporting and assessment (mERA) checklist.
[DOCX File , 22 KB - jmir_v23i5e23350_app5.docx]

References

1. Human papillomavirus (HPV) and cervical cancer. World Health Organization. 2020 Nov 11. URL: https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer [accessed 2020-06-16]
2. Catarino R, Petignat P, Dongui G, Vassilakos P. Cervical cancer screening in developing countries at a crossroad: emerging technologies and policy choices. World J Clin Oncol 2015 Dec 10;6(6):281-290 [FREE Full text] [doi: 10.5306/wjco.v6.i6.281] [Medline: 26677441]
3. Demment MM, Peters K, Dykens JA, Dozier A, Nawaz H, McIntosh S, et al. Developing the evidence base to inform best practice: a scoping study of breast and cervical cancer reviews in low- and middle-income countries. PLoS One 2015 Sep 1;10(9):e0134618 [FREE Full text] [doi: 10.1371/journal.pone.0134618] [Medline: 26325181]
4. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 06;1(2):160-171. [doi: 10.9745/ghsp-d-13-00031]
5. Mehl G, Labrique A. Prioritizing integrated mHealth strategies for universal health coverage. Sci 2014 Sep 12;345(6202):1284-1287. [doi: 10.1126/science.1258926] [Medline: 25214614]
6. Jit M, Brisson M, Portnoy A, Hutubessy R. Cost-effectiveness of female human papillomavirus vaccination in 179 countries: a PRIME modelling study. Lancet Glob Health 2014 Jul;2(7):406-414. [doi: 10.1016/s2214-109x(14)70237-2]
7. Gelband H, Sankaranarayanan R, Gauvreau CL, Horton S, Anderson BO, Bray F, et al. Costs, affordability, and feasibility of an essential package of cancer control interventions in low-income and middle-income countries: key messages from Disease Control Priorities, 3rd edition. Lancet 2016 May;387(10033):2133-2144. [doi: 10.1016/s0140-6736(15)00755-2]
8. Maranga IO, Hampson L, Oliver AW, Gamal A, Gichangi P, Opiyo A, et al. Analysis of factors contributing to the low survival of cervical cancer patients undergoing radiotherapy in Kenya. PLoS One 2013 Oct 30;8(10):e78411 [FREE Full text] [doi: 10.1371/journal.pone.0078411] [Medline: 24205226]
9. Gakidou E, Nordhagen S, Obermeyer Z. Coverage of cervical cancer screening in 57 countries: low average levels and large inequalities. PLoS Med 2008 Jun 17;5(6):e132 [FREE Full text] [doi: 10.1371/journal.pmed.0050132] [Medline: 18563963]
10. World Health Organization. Who Guidelines for Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention. Geneva: World Health Organization; 2013:1-58.

XSL•FO
RenderX

11. Comprehensive Cervical Cancer Control: A Guide to Essential Practice. Second Edition. World Health Organization. 2014. URL: http://apps.who.int/iris/bitstream/handle/10665/144785/9789241548953_eng.pdf [accessed 2021-04-29]
12. Sankaranarayanan R, Budukh AM, Rajkumar R. Effective screening programmes for cervical cancer in low- and middle-income developing countries. Bull World Health Organ 2001;79(10):954-962 [FREE Full text] [Medline: 11693978]
13. World Health Organization. Global action plan for the prevention and control of NCDs 2013-2020. In: Noncommunicable Diseases and Mental Health. Geneva: World Health Organization; 2013:1-55.
14. WHO, ITU. Be He@lthy, Be Mobile Annual Report 2016. In: Noncommunicable Diseases and Their Risk Factors. Geneva: World Health Organization; 2016.
15. Gidman W, Coomber R. Contested space in the pharmacy: public attitudes to pharmacy harm reduction services in the West of Scotland. Res Social Adm Pharm 2014 May;10(3):576-587. [doi: 10.1016/j.sapharm.2013.07.006] [Medline: 24021859]
16. Im H, Castro CM, Shao H, Liong M, Song J, Pathania D, et al. Digital diffraction analysis enables low-cost molecular diagnostics on a smartphone. Proc Natl Acad Sci U S A 2015 May 05;112(18):5613-5618 [FREE Full text] [doi: 10.1073/pnas.1501815112] [Medline: 25870273]
17. 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 Aug 27;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
18. Sahin D, Hacisalihoglu UP, Kirimlioglu SH. Telecytology: is it possible with smartphone images? Diagn Cytopathol 2018 Jan 08;46(1):40-46. [doi: 10.1002/dc.23851] [Medline: 29115040]
19. Tanaka Y, Ueda Y, Okazawa A, Kakuda M, Matsuzaki S, Kobayashi E, et al. 'Smartscopy' as an alternative device for cervical cancer screening: a pilot study. BMJ Innov 2017 Apr 30;3(2):123-126 [FREE Full text] [doi: 10.1136/bmjinnov-2016-000174] [Medline: 28794895]
20. DiCarlo J, Gopakumar S, Dhillon P, Krishnan S. Adoption of information and communication technologies for early detection of breast and cervical cancers in low- and middle-income countries. J Glob Oncol 2016 Aug;2(4):222-234 [FREE Full text] [doi: 10.1200/JGO.2015.002063] [Medline: 28717705]
21. 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]
22. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
23. Francis DB, Cates JR, Wagner KP, Zola T, Fitter JE, Coyne-Beasley T. Communication technologies to improve HPV vaccination initiation and completion: a systematic review. Patient Educ Couns 2017 Jul;100(7):1280-1286. [doi: 10.1016/j.pec.2017.02.004] [Medline: 28209248]
24. Smulian EA, Mitchell KR, Stokley S. Interventions to increase HPV vaccination coverage: a systematic review. Hum Vaccin Immunother 2016 Jun 02;12(6):1566-1588 [FREE Full text] [doi: 10.1080/21645515.2015.1125055] [Medline: 26838959]
25. World bank country and lending groups. World Bank Group. 2018. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2021-03-12]
26. Noyes JH, Booth A, Harris J, Harden A, Popay J, Pearson A, Noyes. Chapter 5: Extracting qualitative evidence. Cochrane Qualitative & Implementation Methods Group. 2010. URL: http://cqrmg.cochrane.org/supplemental-handbook-guidance [accessed 2021-03-12]
27. Quality assessment tool for quantitative studies. Effective Public Health Practice Project. 2010. URL: https://link.springer.com/content/pdf/bbm%3A978-3-319-17284-2%2F1.pdf [accessed 2021-03-12]
28. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
29. Torgerson C. Systematic Reviews. London, United Kingdom: A&C Black; 2003:1-102.
30. Asgary R, Adongo PB, Nwameme A, Cole HV, Maya E, Liu M, et al. mHealth to train community health nurses in visual inspection with acetic acid for cervical cancer screening in Ghana. J Low Genit Tract Dis 2016;20(3):239-242. [doi: 10.1097/lgt.0000000000000207]
31. Asgary R, Cole H, Adongo P, Nwameme A, Maya E, Adu-Amankwah A, et al. Acceptability and implementation challenges of smartphone-based training of community health nurses for visual inspection with acetic acid in Ghana: mHealth and cervical cancer screening. BMJ Open 2019 Jul 16;9(7):e030528 [FREE Full text] [doi: 10.1136/bmjopen-2019-030528] [Medline: 31315879]
32. Bhatt S, Isaac R, Finkel M, Evans J, Grant L, Paul B, et al. Mobile technology and cancer screening: lessons from rural India. J Glob Health 2018 Dec;8(2):020421 [FREE Full text] [doi: 10.7189/jogh.08.020421] [Medline: 30603075]

XSL•FO
RenderX

33. Devi S, Dasila P, Suryavanshi MR. Effect of educational intervention followed by reminders regarding participation of women in the cervical screening programme. Ind J Pub Health Res & Develop 2018;9(9):369. [doi: 10.5958/0976-5506.2018.01026.4]
34. Caster MM, Norris AH, Butao C, Reese PC, Chemey E, Phuka J, et al. Assessing the acceptability, feasibility, and effectiveness of a tablet-based cervical cancer educational intervention. J Cancer Educ 2017 Mar 5;32(1):35-42 [FREE Full text] [doi: 10.1007/s13187-015-0953-6] [Medline: 26637473]
35. Catarino R, Vassilakos P, Scaringella S, Undurraga-Malinverno M, Meyer-Hamme U, Ricard-Gauthier D, et al. Smartphone use for cervical cancer screening in low-resource countries: a pilot study conducted in Madagascar. PLoS One 2015 Jul 29;10(7):e0134309 [FREE Full text] [doi: 10.1371/journal.pone.0134309] [Medline: 26222772]
36. Gallay C, Girardet A, Viviano M, Catarino R, Benski A, Tran PL, et al. Cervical cancer screening in low-resource settings: a smartphone image application as an alternative to colposcopy. Int J Women's Health 2017 Jun;Volume 9:455-461. [doi: 10.2147/ijwh.s136351]
37. 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]
38. 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 22;5(1):28-34 [FREE Full text] [doi: 10.1136/bmjinnov-2018-000276] [Medline: 31645991]
39. Huchko MJ, Saduma I, Blat C, Oketch S, Bukusi EA. How providing cervical cancer screening results via cell phone affects patient follow-up rates in Western Kenya. J Glob Oncol 2019 Dec(5):1-8. [doi: 10.1200/jgo.18.00264]
40. 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]
41. Lima TM, Nicolau AI, Carvalho FH, Vasconcelos CT, Aquino PD, Pinheiro AK. Telephone interventions for adherence to colpocytological examination. Rev Lat Am Enfermagem 2017 Feb 06;25:e2844 [FREE Full text] [doi: 10.1590/1518-8345.1683.2844] [Medline: 28177055]
42. Littman-Quinn R, Mibenge C, Antwi C, Chandra A, Kovarik CL. Implementation of m-health applications in Botswana: telemedicine and education on mobile devices in a low resource setting. J Telemed Telecare 2013 Feb 01;19(2):120-125. [doi: 10.1177/1357633x12474746] [Medline: 23454821]
43. Ndlovu K, Littman-Quinn R, Park E, Dikai Z, Kovarik CL. Scaling up a mobile telemedicine solution in Botswana: keys to sustainability. Front Public Health 2014 Dec 11;2:275 [FREE Full text] [doi: 10.3389/fpubh.2014.00275] [Medline: 25566520]
44. Parham GP, Mwanahamuntu M, Pfaendler K, Sahasrabuddhe VV, Myung D, Mkumba G, et al. eC3--a modern telecommunications matrix for cervical cancer prevention in Zambia. J Low Genit Tract Dis 2010 Jul;14(3):167-173 [FREE Full text] [doi: 10.1097/LGT.0b013e3181cd6d5e] [Medline: 20592550]
45. Peterson CW, Rose D, Mink J, Levitz D. Real-time monitoring and evaluation of a visual-based cervical cancer screening program using a decision support job aid. Diagnostics (Basel) 2016 May 16;6(2):20 [FREE Full text] [doi: 10.3390/diagnostics6020020] [Medline: 27196932]
46. Quinley KE, Gormley RH, Ratcliffe SJ, Shih T, Szep Z, Steiner A, et al. Use of mobile telemedicine for cervical cancer screening. J Telemed Telecare 2011 May 06;17(4):203-209 [FREE Full text] [doi: 10.1258/jtt.2011.101008] [Medline: 21551217]
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 Oct 30;14(10):5901-5904 [FREE Full text] [doi: 10.7314/apjcp.2013.14.10.5901] [Medline: 24289597]
48. Rashid RM, Dahlui M. Study protocol for the most effective recall method in a cervical cancer screening program in Klang, Malaysia. Asian Pac J Cancer Prev 2013 Oct 30;14(10):5867-5870 [FREE Full text] [doi: 10.7314/apjcp.2013.14.10.5867] [Medline: 24289591]
49. Rashid RM, Ramli S, John J, Dahlui M. Cost effective analysis of recall methods for cervical cancer screening in Selangor--results from a prospective randomized controlled trial. Asian Pac J Cancer Prev 2014 Jul 15;15(13):5143-5147 [FREE Full text] [doi: 10.7314/apjcp.2014.15.13.5143] [Medline: 25040965]
50. Ricard-Gauthier D, Wisniak A, Catarino R, van Rossum AF, Meyer-Hamme U, Negulescu R, et al. Use of smartphones as adjuvant tools for cervical cancer screening in low-resource settings. J Low Genit Tract Dis 2015 Oct;19(4):295-300. [doi: 10.1097/LGT.0000000000000136] [Medline: 26247260]
51. Sharma D, Rohilla L, Bagga R, Srinivasan R, Jindal HA, Sharma N, et al. Feasibility of implementing cervical cancer screening program using smartphone imaging as a training aid for nurses in rural India. Public Health Nurs 2018 Nov 28;35(6):526-533. [doi: 10.1111/phn.12517] [Medline: 29806745]

XSL•FO
RenderX

52. Swanson M, Ibrahim S, Blat C, Oketch S, Olwanda E, Maloba M, et al. Evaluating a community-based cervical cancer screening strategy in Western Kenya: a descriptive study. BMC Womens Health 2018 Jul 03;18(1):116 [FREE Full text] [doi: 10.1186/s12905-018-0586-0] [Medline: 29970063]
53. Taghavi K, Banerjee D, Mandal R, Kallner HK, Thorsell M, Friis T, et al. Colposcopy telemedicine: live versus static swede score and accuracy in detecting CIN2+, a cross-sectional pilot study. BMC Womens Health 2018 Jun 11;18(1):89 [FREE Full text] [doi: 10.1186/s12905-018-0569-1] [Medline: 29890991]
54. Tran PL, Benski C, Viviano M, Petignat P, Combescure C, Jinoro J, et al. Performance of smartphone-based digital images for cervical cancer screening in a low-resource context. Int J Technol Assess Health Care 2018 Jan;34(3):337-342. [doi: 10.1017/S0266462318000260] [Medline: 29921339]
55. Urner E, Delavy M, Catarino R, Viviano M, Meyer-Hamme U, Benski A, et al. A smartphone-based approach for triage of human papillomavirus-positive sub-saharan african women: a prospective study. JMIR Mhealth Uhealth 2017 May 29;5(5):e72 [FREE Full text] [doi: 10.2196/mhealth.6697] [Medline: 28554879]
56. Yeates KE, Sleeth J, Hopman W, Ginsburg O, Heus K, Andrews L, et al. Evaluation of a smartphone-based training strategy among health care workers screening for cervical cancer in Northern Tanzania: the Kilimanjaro method. J Glob Oncol 2016 Dec;2(6):356-364. [doi: 10.1200/jgo.2015.001768]
57. Yeates K, Erwin E, Mtema Z, Magoti F, Nkumbugwa S, Yuma S, et al. Smartphone-enhanced training, qa, monitoring, and evaluation of a platform for secondary prevention of cervical cancer: opportunities and challenges to implementation in Tanzania. JCO Glob Oncol 2020 Nov(6):1114-1123. [doi: 10.1200/go.20.00124]
58. 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. [doi: 10.31557/apjcp.2020.21.3.853]
59. 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]
60. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014 Jun 13;10(1):49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]

Abbreviations

HPV: human papillomavirus
LMIC: low- and middle-income country
mERA: mobile health evidence reporting and assessment
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
VIA: visual inspection with acetic acid

Edited by G Eysenbach; submitted 17.08.20; peer-reviewed by K Morley, W Evans; comments to author 11.09.20; revised version received 01.01.21; accepted 13.01.21; published 27.05.21.

Please cite as:
Rossman AH, Reid HW, Pieters MM, Mizelle C, von Isenburg M, Ramanujam N, Huchko MJ, Vasudevan L
Digital Health Strategies for Cervical Cancer Control in Low- and Middle-Income Countries: Systematic Review of Current Implementations and Gaps in Research
J Med Internet Res 2021;23(5):e23350
URL: https://www.jmir.org/2021/5/e23350
doi:10.2196/23350
PMID:34042592

©Andrea H Rossman, Hadley W Reid, Michelle M Pieters, Cecelia Mizelle, Megan von Isenburg, Nimmi Ramanujam, Megan J Huchko, Lavanya Vasudevan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

XSL•FO
RenderX

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 Social Media for Health Research Purposes: Scoping Review

Charline Bour[1], MSc; Adrian Ahne[2,3], MSc; Susanne Schmitz[4], PhD; Camille Perchoux[5], PhD; Coralie Dessenne[1]; Guy Fagherazzi[1], MSc, PhD

[1]Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

[2]Inserm U1018, Center for Research in Epidemiology and Population Health (CESP), Paris Saclay University, Villejuif, France

[3]Epiconcept, Paris, France

[4]Competence Centre for Methodology and Statistics, Luxembourg Institute of Health, Strassen, Luxembourg

[5]Luxembourg Institute of Socio-Economic Research, Esch/Alzette, Luxembourg

Corresponding Author:
Guy Fagherazzi, MSc, PhD
Department of Population Health
Luxembourg Institute of Health
1 A-B Rue Thomas Edison
Strassen, 1445
Luxembourg
Phone: 352 26970 457
Fax: 352 26970 719
Email: guy.fagherazzi@lih.lu

Related Article:

Comment in: https://www.jmir.org/2022/1/e35286

Abstract

Background: As social media are increasingly used worldwide, more and more scientists are relying on them for their health-related projects. However, social media features, methodologies, and ethical issues are unclear so far because, to our knowledge, there has been no overview of this relatively young field of research.

Objective: This scoping review aimed to provide an evidence map of the different uses of social media for health research purposes, their fields of application, and their analysis methods.

Methods: We followed the scoping review methodologies developed by Arksey and O'Malley and the Joanna Briggs Institute. After developing search strategies based on keywords (eg, social media, health research), comprehensive searches were conducted in the PubMed/MEDLINE and Web of Science databases. We limited the search strategies to documents written in English and published between January 1, 2005, and April 9, 2020. After removing duplicates, articles were screened at the title and abstract level and at the full text level by two independent reviewers. One reviewer extracted data, which were descriptively analyzed to map the available evidence.

Results: After screening 1237 titles and abstracts and 407 full texts, 268 unique papers were included, dating from 2009 to 2020 with an average annual growth rate of 32.71% for the 2009-2019 period. Studies mainly came from the Americas (173/268, 64.6%, including 151 from the United States). Articles used machine learning or data mining techniques (60/268) to analyze the data, discussed opportunities and limitations of the use of social media for research (59/268), assessed the feasibility of recruitment strategies (45/268), or discussed ethical issues (16/268). Communicable (eg, influenza, 40/268) and then chronic (eg, cancer, 24/268) diseases were the two main areas of interest.

Conclusions: Since their early days, social media have been recognized as resources with high potential for health research purposes, yet the field is still suffering from strong heterogeneity in the methodologies used, which prevents the research from being compared and generalized. For the field to be fully recognized as a valid, complementary approach to more traditional health research study designs, there is now a need for more guidance by types of applications of social media for health research, both from a methodological and an ethical perspective.

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-040671

(J Med Internet Res 2021;23(5):e25736) doi:10.2196/25736

KEYWORDS

social media; public health; epidemiology; research; health; medical; social networking; infodemiology; eHealth; text mining

Introduction

Social Media Background

Social media (SM) refer to new forms of media that involve interactions between users [1] in personal (eg, Facebook) or more professional (eg, LinkedIn) ways. In 2010 in the United States, 80% of adults used the internet to search for health-related information, and 11% of SM users posted comments, queries, or information about health or medical content [2]. Every user activity on the internet generates a unique digital footprint that can be collected for health research [3]. However, SM are not only used in a personal way. Indeed, academics are also increasingly using SM to share their work and disseminate their findings [4].

Opportunities for Health Research

Since the creation of SM in 2004-2005 and with 3.81 billion active social media users in April 2020 [5], concepts like infodemiology and infoveillance have emerged. The term "infodemiology" refers to the science of using the internet to improve public health, while "infoveillance" refers to the science of syndromic surveillance using the internet [6]. These opportunities have been seized through the years in order to create new methodologies for health research to cope with the issues raised by traditional methods (eg, difficulty of recruitment [7]).

Scoping Review Contextualization

Previous scoping and systematic reviews have already been published about the different uses of SM for health research. However, they were either focusing on a specific type of SM (eg, blogs [8]), on a specific field of health research (eg, child maltreatment [9]), or on a specific methodology (eg, recruitment of study participants [10,11]). Other reviews discussed the overall use of SM for health research [12,13] but did not provide any insights on the analysis techniques or the ethical issues. Besides, the COVID-19 pandemic has sped things up and pushed research to be done online, leveraging existing data for disease surveillance purposes, which makes the present work particularly timely and needed for better structuration of the field [14]. The research field on social media and health is relatively young and therefore lacks structures and guidelines. In the light of the above, it seemed important to map the different uses of social media for health research. Our work will directly contribute to the general effort of acknowledging the potential of this research field and will help to identify the main limitations to tackle in the future.

Review Questions

The overall research questions were as follows: (1) How have SM modified or complemented traditional health research? (2) What are the different fields of application of this approach? (3) What are the different methodologies for SM data analysis?

Methods

Overview

This scoping review followed the methodological framework introduced by Arskey and O'Malley in 2005 [15] and the methodology manual published by the Joanna Briggs Institute for scoping reviews [16]. It is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Review (PRISMA-ScR) guidelines [17]. The methods have been previously detailed in a research protocol [18].

Search Strategy

An initial literature search was first manually conducted on PubMed/MEDLINE to identify the health research fields in which SM are mostly used and developed. We searched for the term "social media" in the Medical Subject Headings (MeSH) Terms (words or phrases selected to represent particular biomedical concepts) as it gathers all papers discussing the use of at least one example of social media. For instance, this MeSH Term also includes articles that mention Facebook or Twitter without referring directly to "social media." We considered the term "health research" as all kinds of research performed to learn more about human health, prevent or treat disease, test ideas, improve treatments, and answer questions. Then, the literature search was performed through PubMed/MEDLINE and Web of Science. The search strategy, highlighted in Textbox 1, included two sets of search terms: (1) one linked with SM (eg, social media) and (2) one linked with research (eg, health research, biomedical research). In order to capture the evolution of SM uses for health research over the years, databases were searched between January 1, 2005, and April 9, 2020. The term "social network" was also searched, as it is often misused as a synonym of SM. An additional list of 5 relevant articles [19-23] was manually searched to identify any other potentially relevant articles not yet captured. These articles were chosen in order to retrieve more articles about infodemiology, ethical issues, or the use of SM data. A snowball searching technique was adopted with these 5 articles in which citations within articles were searched and kept if relevant to the review.

XSL•FO
RenderX

Textbox 1. Inclusion criteria, exclusion criteria, and search strings.

Inclusion criteria

- written in English
- published between January 1, 2005, and April 9, 2020
- dealt with the use of social media by researchers

Exclusion criteria

- not about health research
- not related to social media (eg, social network analysis)
- not about human subjects
- no relevant information (eg, methodology) about the use of social media for health research
- no relevant characteristics of social media

Search string in PubMed

(((“Social Media”[MH]) OR (“Social Media”[TW])) AND ((“Biomedical research”[MH]) OR (“Medical research”[TW] OR “Biomedical research”[TW]) OR (“Health research”[TW] OR “Health services research”[TW]))) OR (((“Social networking”[MH]) OR (“Social network”[TW] OR “Social networks”[TW] OR “Social networking”[TW])) AND ((“Biomedical research”[MH]) OR (“Medical research”[TW] OR “Biomedical research”[TW]) OR (“Health research”[TW] OR “Health services research”[TW]))) Filters: Journal Article; Publication date from 2005/01/01 to 2020/04/09; Humans; English

Search string in Web of Science

(TS=“Social Media” OR TS= “Social networking” OR TS= “Social network” OR TS= “Social networks”) AND (TS=“Biomedical research” OR TS=“Medical research” OR TS=“Health research” OR TS=“Health services research”) AND (PY=(2005-2020)) AND (LANGUAGE: (English)) Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

Eligibility Criteria

This review was guided by the “Population, Concept, Context” framework suggested by the Joanna Briggs Institute [24]. We did not have any restriction about the population; we took any relevant publications regardless of the age, the origin, or the gender of the studied populations. The concept was the use of social media and the context was health research. The eligibility criteria were any journal article that described the use of social media platforms or social media data for health or medical research purposes. We excluded articles that were not directly related to research from our review, such as those on the use of social media among patients, patient associations or communities, organizations, or health care professionals for their day-to-day practice. Grey literature and studies about nonhuman subjects were excluded as well. Documents related to the mining of social media data to detect prescription drug misuse and abuse as well as those related to the use of machine learning methodologies to analyze data were eligible for inclusion. We included full texts that reported on at least one of the following outcomes: (1) SM data analysis, (2) recruitment through SM, (3) methodology for SM research, and (4) ethical issues of using SM for health research. Only English-language articles were retained. The inclusion and exclusion criteria and the search strings are summarized in Textbox 1.

Study Selection Process

A 2-step screening was performed after duplicate removal. First, titles and abstracts were screened in order to define the eligibility of each article. Publications with title or abstract not meeting the eligibility criteria were excluded. Then, the full texts having passed the first step were screened, and only articles meeting the eligibility criteria were kept. All screening levels were conducted with CADIMA [25], a free web tool to facilitate the conduct and documentation of literature reviews [26]. Two reviewers screened articles (GF, CB) independently, and consistency checks were performed thanks to CADIMA.

Data Extraction

Data were abstracted on (1) the country of origin, (2) the aims of the study (eg, to map ethical issues when using SM for health research), (3) the type of study (eg, recruitment feasibility assessment), (4) the research field (eg, mental health research), (5) the studied population (eg, adolescents), (6) the type of SM (eg, Facebook), (7) the methodology (eg, paid advertisement), (8) the outcomes of the study (eg, efficiency of recruitment via SM), and (9) the key findings for our scoping review (eg, possibility to recruit on SM). Data were extracted and cleaned by a first reviewer (CB), then verified and approved by a second reviewer (GF).

Methodological Quality Appraisal

Because this is a scoping review, we did not appraise methodological quality or risk of bias of the included articles.

Analysis and Presentation of Results

We conducted a descriptive analysis of the characteristics of the included literature. We described the included articles according to the journal of publication, publication date, country of origin (location of the corresponding author), Altmetric score (automatically calculated weighted count of all of the attention a research output has received) [27], type of SM, type of population, and type of disease studied. We decided to focus on Altmetric score rather than citation counts; as the SM

XSL•FO
RenderX

research field is still relatively young, traditional citation counts provide a quite conservative approach of a paper's "influence" that is influenced by the size of the research community working on the topic. Thus, Altmetric might be less influenced by the relatively "young" aspect of this research field by giving weight to other dimensions (record of dissemination, influence, impact). All these measures are more nuanced than citation counts alone are able to be [28,29]. However, Altmetric scores also have some limitations, as they do not take comparability across journals and platforms into account, and this system can be gamed [30,31].

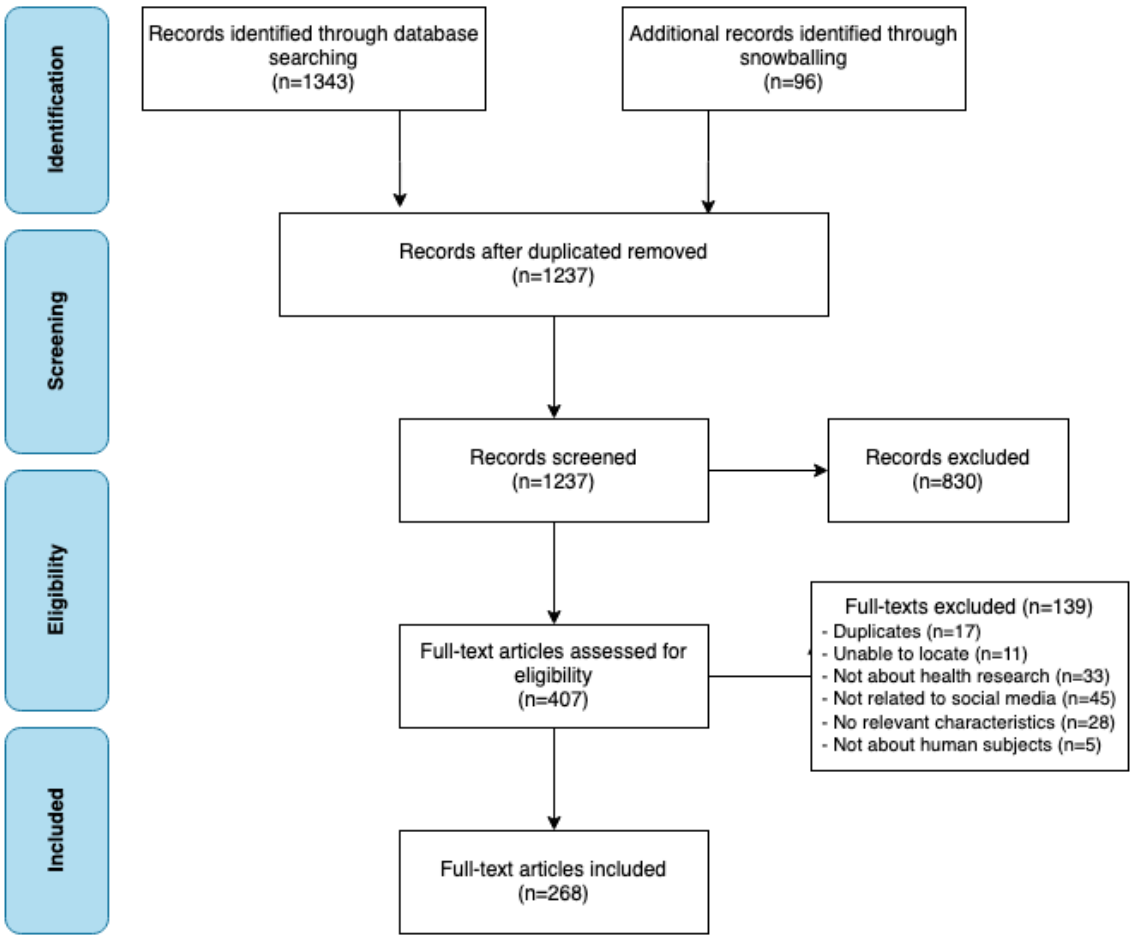
We categorized the diseases in 7 categories: (1) chronic diseases (eg, diabetes), (2) communicable diseases (eg, influenza), (3) alcohol/smoking (eg, vaping), (4) mental health (eg, depression), (5) lifestyle (eg, nutrition outcomes), (6) drug/medication (eg, drug use disorder), and (7) other (eg, child maltreatment). Descriptive statistics and corresponding plots were computed (n, means, frequencies) with R (version 3.6.3; R Foundation for Statistical Computing).

Results

Search Results

The initial search conducted in April 2020 revealed 1343 results. An additional 96 articles were retrieved through a snowballing technique based on 5 relevant articles [16-20]. This resulted in a total of 1439 articles, and duplicates (n=202) were removed. Then, 1237 titles and abstracts were screened, which led to the exclusion of 830 articles. Overall, 407 studies were included to screen as full - text papers, of which 139 were excluded. The main reasons for exclusion were that the study (1) did not contain relevant characteristics of SM for health research (n=28), (2) did not relate to SM (n=45), or (3) was not about health research (n=33). 268 studies were included in the analyses. Figure 1 shows the flow diagram of the article selection. Lastly, Multimedia Appendix 1 displays the characteristics of the 268 included studies (author or authors, year of publication, country, title, aim of the study, type of social media, studied population and disease).

Figure 1. Flow diagram of the included studies.



Distribution of Studies

In all, we included 268 unique records from 155 different journals. Table 1 displays the 10 most common journals in which the included studies were published: 55 (20.5%) articles were published in the *Journal of Medical Internet Research* or sister journals *JMIR Research Protocols* and *JMIR Public Health and Surveillance*. *PLoS ONE* is the second most common journal with 10 (3.7%) articles.

XSL•FO
RenderX

Table 1. Top 10 most common journals publishing work using social media for health research purposes.

Name of the journal	Articles, n (%)
Journal of Medical Internet Research	39 (14.6)
PLoS ONE	10 (3.7)
JMIR Research Protocols	9 (3.4)
JMIR Public Health and Surveillance	7 (2.6)
American Journal of Public Health	5 (1.9)
The American Journal of Bioethics	5 (1.9)
BMC Medical Informatics and Decision Making	4 (1.5)
International Journal of Environmental Research and Public Health	4 (1.5)
PLoS Computational Biology	4 (1.5)
Digital Health	3 (1.1)

A total of 1025 authors took part in the writing of the included studies. Figure S1 in Multimedia Appendix 2 provides the coauthorship network of all these authors. The largest set of connected authors included 57 authors and shown in Figure S2 in Multimedia Appendix 2).

Even though our research date range was from 2005 to 2020, none of the 268 included articles are dated before 2009. In Table 2, it can be seen that the number of publications is growing through the years, corresponding to an average annual growth rate of 32.7% for the 2009-2019 period. This suggests that the field of health research supplemented by SM has gained interest for the last 11 years. Earlier studies concentrated more on the use of SM for health research in general and the opportunities for the study of communicable diseases. The most recent studies more frequently included recruitment strategies and methodologies. Table 3 displays the distribution of articles by the continent of publication. Most articles were from the Americas (173/268, 64.6%, including 151/173, 87.3% from the United States), 18.7% were from Europe (50/268), 11.6% were from Oceania (31/268), 4.9% were from Asia (13/268), and 0.4% were from Africa (1/268).

Table 2. Distribution of publications by year of publication.

Year	Publications, n
2009	2
2010	3
2011	7
2012	5
2013	20
2014	26
2015	34
2016	37
2017	42
2018	36
2019	45
2020 (Jan 1–Apr 9)	11

Table 3. Distribution of publications by geographic location (as assessed by the location of the corresponding author).

Geographic location	Publications, n
Africa	1
The Americas	173
Asia	13
Europe	50
Oceania	31

XSL•FO
RenderX

Social Media

Among all the retrieved articles, 57.8% (155/268) used or described at least one specific type of SM. From these articles, as can be seen in Table 4, 42.6% (66/155) were based on Twitter, 34.2% (53/155) on Facebook, and 11.0% (17/155) on several SM (eg, combining Facebook, Instagram, and Snapchat [32]). The remaining 12.3% (19/155) were distributed between Instagram, Reddit, forums, blogs, Weibo, and YouTube.

Table 4. Distribution of publications by social media (N=155).

Type of social media	Publications, n
Blogs	2
Facebook	53
Forums	3
Instagram	5
Reddit	5
Several types	17
Twitter	66
Weibo	2
YouTube	2

Focused Populations

A total of 80.2% (215/268) of included articles did not focus on any specific population. In articles that studied a specific subpopulation (n=53), youth was the most common one (34/53, 64%), followed by women (7/53, 13%), families (5/53, 9%), men (1/53, 2%), and other (6/53, 11%), as shown in Table 5. The "Other" category gathered adults (2/6), Chinese migrants (1/6), elderly people (1/6), emergency nurses (1/6), and researchers (1/6).

Table 5. Distribution of publications per studied population (N=53).

Studied population	Publications, n
Youth	34
Women	7
Families	5
Men	1
Other	6

Domain of Health Research

In addition, 45.1% (121/268) of publications dealt with a specific disease or condition (the remaining articles usually dealt with the use of SM for health research in general or with methodology). Indeed, as shown in Table 6, 33.1% (40/121) of articles studied communicable diseases, 19.8% (24/121) studied chronic diseases, 15.7% (19/121) studied lifestyle (eg, nutrition outcomes), 13.2% (16/121) studied other conditions (eg, drug use disorder), 9.9% (12/121) studied alcohol/smoking (eg, vaping), and 8.3% (10/121) studied mental health (eg, depression).

Table 6. Distribution of publications by studied disease type (N=121).

Studied disease type	Publications, n
Alcohol/smoking	12
Chronic diseases	24
Communicable diseases	40
Lifestyle	19
Mental health	10
Other	16

Communicable Diseases

Among articles that discussed communicable diseases, influenza was the primary studied disease (18/40, 45%), followed by HIV (8/40, 20%) and human papillomavirus (3/40, 8%).

Chronic Diseases

Among articles that discussed chronic diseases, a quarter studied cancer (6/24, 25%), followed by diabetes (5/24, 21%),

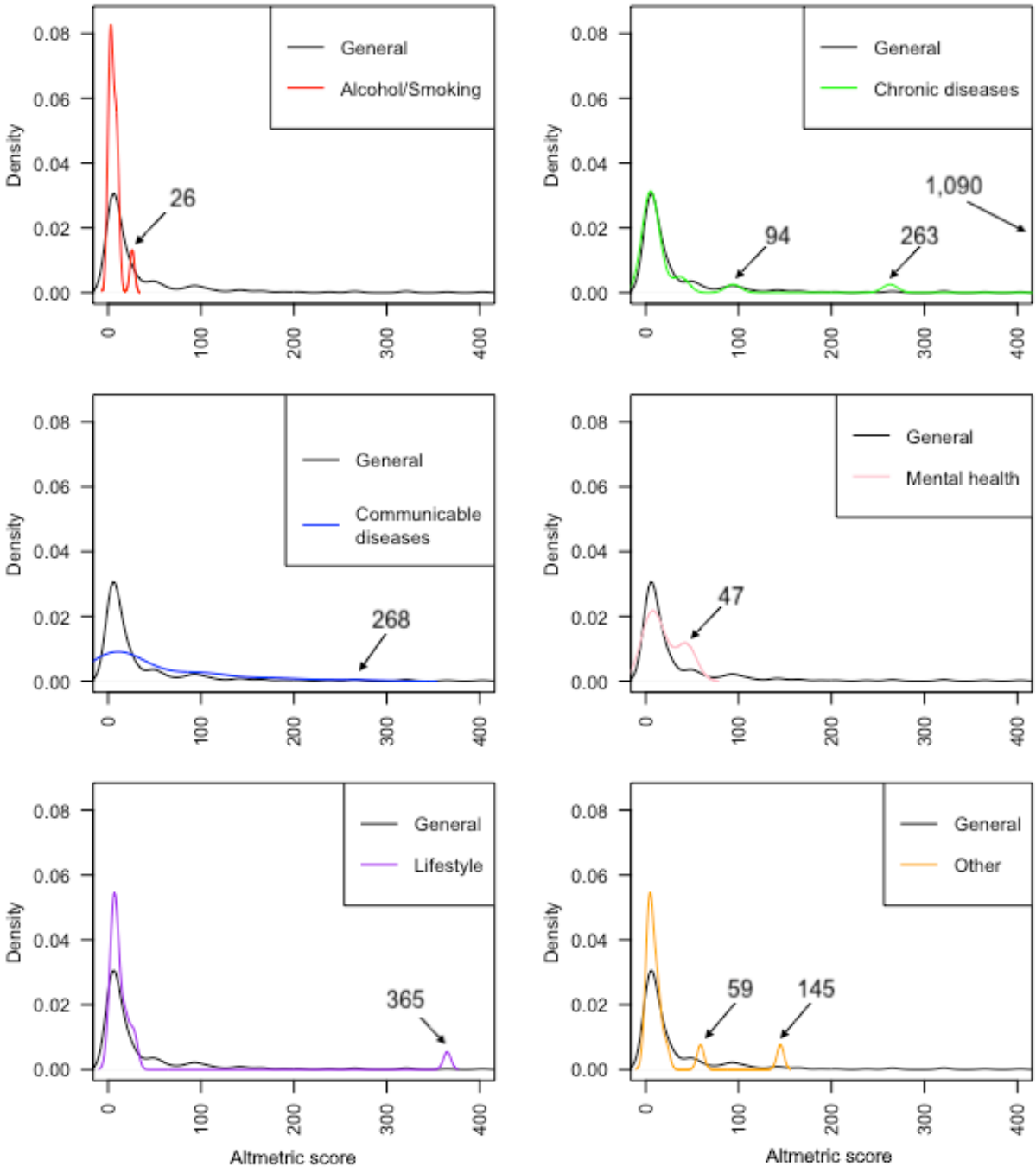
XSL•FO
RenderX

cardiovascular diseases (eg, congenital heart disease, 3/24, 12%), and obesity (2/24, 8%).

Dissemination

As highlighted in Figure 2, some papers stood out and could be considered important in the recent field of health research on social media.

Figure 2. Distribution of Altmetric scores by health research area. Corresponding publications to the indicated Altmetric scores: Alcohol/Smoking: 26 [33]; Chronic diseases: 94 [34], 263 [35], 1090 [36]; Communicable diseases: 268 [37]; Mental health: 47 [38]; Lifestyle: 365 [39]; Other: 59 [40], 145 [41]. General corresponds to the Altmetric scores of all studies.



Type of Studies

Among all included studies, 22.4% (60/268) described the use of machine learning and data mining techniques, 22.0% (59/268) discussed the opportunities and limitations of the use of SM for research, 16.8% (45/268) assessed the feasibility of recruitment strategies on SM, 6.0% (16/268) discussed the ethical issues when using SM for health research, 5.2% (14/268) gave methodologies for health research, and 4.9% (13/268) illustrated the use of SM for dissemination. Guidelines for recruitment (9/268, 3.4%), interventions of prevention (6/268, 2.2%), crowdfunding (4/268, 1.5%), sentiment analysis (4/268, 1.5%), data anonymization (2/268, 0.7%), and crowdsourcing (2/268, 0.7%) were also considered.

Machine Learning and Other Techniques

Machine learning techniques included text mining (17/60, 28%), natural language processing (15/60, 25%), data mining (12/60,

XSL•FO
RenderX

20%), classification (10/60, 17%), topic modelling (4/60, 7%), deep learning (1/60, 2%), and social network analysis (1/60, 2%). In particular, support vector machine (17/60, 28%), logistic regression (11/60, 18%), latent Dirichlet allocation (5/60, 8%), convolutional neural network (5/60, 8%), random forests (4/60, 7%), decision trees (4/60, 7%) and n-grams (64/60, 7%) are the most used models. Stacked linear regression, Bayesian network algorithm, nonnegative matrix factorization, stochastic gradient, learning vector quantization, and recurrent neural networks represent 2% (1/60) each. Lastly, these techniques were mostly used for data coming from Twitter (38/60, 63%) and Reddit (3/60, 5%).

Recruitment Strategies

Studies assessing recruitment strategies' feasibility applied paid advertisement (36/45, 80%), free advertisement (eg, posting in relevant Facebook groups [42]) (6/45, 13%), and the combination of both advertisements (3/45, 7%). Paid recruitment strategies included designing the ad, targeting the right audience with Facebook Ads Manager and measuring the impacts with Facebook Analytics [43]. Moreover, 64% (29/45) of studies considered SM recruitment to be effective (time-effective and efficient to recruit populations). Paid advertisement was evaluated as cost-effective in 83% (30/36) of studies and too costly in 6% (2/36). We found out that 80% (36/45) of recruitment was carried out on Facebook, 9% (4/45) on both Facebook and Twitter, and 9% (4/45) on more than two types of SM (eg, Facebook, Twitter, Craigslist, Tumblr, LinkedIn [44]). Lastly, a third of recruitment strategies included providing incentives to participants (eg, gift cards).

Ethical Issues

Ethical issues were usually mentioned but not investigated in detail. When the article focused on ethical issues (n=16), the main ethical issues raised were getting consent of online users (15/16, 94%), protecting the privacy of users (14/16, 88%), preserving confidentiality (9/16, 56%), potential harms to participants (9/16, 56%), preserving of anonymity (8/16, 50%), securing data (7/16, 44%), transparency of the research (7/16, 44%), application of guidelines (7/16, 44%), representativeness and self-selection bias (5/16, 31%), and the risk of double accounts (2/16, 12%).

Discussion

Principal Findings

The overarching aim of this review was to scope the literature for evidence on the use of SM for health research. We were able to include 268 studies. Most of the included articles in this scoping review are dated from 2013 onwards, which is consistent with the worldwide growth of SM use over the last decade [45]. We identified three main SM used for health research: Twitter, Facebook, and Instagram, the most popular platforms in 2020 [46]. The most studied populations are young adults and adolescents. This could be related to the elevated proportion of young people active on SM. In 2018 in the United States, 51% of teens were on Facebook, 69% on Snapchat, 72% on Instagram, and 85% on YouTube; thus, SM seems to have great potential to focus on the young generations [47]. The majority of the included works focused on both communicable and chronic diseases. The field of SM research is still very young, and this can be seen in the impact that publications have (via the Altmetric score), with the exception of 9 articles. However, it is set to evolve rapidly, and it will be necessary to follow the evolution of the Altmetric scores of the field in the coming years to identify the new major articles.

The fields of application of SM in health research are broad and constantly evolving: as earlier studies concentrate on the study of communicable diseases, most recent studies include recruitment strategies and data collection for infoveillance. First, SM can be used to complement traditional methods. Traditional procedures can meet several limitations. When recruiting a specific population, traditional methods (eg, fliers, advertising) can be expensive or limited in reach [7,48-50]. Complementing them with SM advertisements can cope with these limitations. Second, SM alone show high potential. Studies have concluded that SM paid advertisements can be an efficient and cost-effective tool to recruit [11,51-56]. SM appear not only to facilitate and complement traditional recruitment strategies to reach specific populations but to be efficient as well when used alone [52,57-59], especially to reduce time constraints or to target a large population [60]. Particularly, Facebook can be used to recruit participants of all ages and allows researchers to obtain participant samples similarly representative to those recruited via traditional recruitment methods [11]. Facebook, together with Facebook Ads Manager and Facebook Analytics, are particularly useful to develop and adjust such strategies. Traditional disease surveillance, population surveillance, and epidemiology methodologies can be improved by SM [21,50,61]. Pharmacovigilance and the detection of adverse drug reactions on SM proved to be efficient and to reduce time between the online report of an incident and its discovery [62-64]. As the number of SM users is increasing, generated data, or "big data," is expanding. Such data can be collected and studied to improve disease and public health surveillance [65-67] to forecast diseases [68] or to improve research in a medical field [69,70]. Along with big data growth, machine learning and data mining techniques such as text mining and natural language processing are constantly evolving and are thus increasingly used in the field of public health research based on SM [71-73]. These techniques can be particularly interesting to analyze social media data and, for instance, to develop sentiment or topic analysis among a specific population [19,74] or to predict epidemics [75]. Twitter is mainly used for such work because Twitter developed a streaming application programming interface. This is a free application that allows easy access to 1% of all Twitter data in real time, filtered by specific criteria (eg, keywords) [76,77]. Lastly, SM can be directly used by health researchers to support prevention interventions to raise awareness and engage populations [78] and to crowdfund by promoting their research on SM. Indeed, crowdfunding can be eased by establishing professional contacts through SM and sharing campaigns [79].

The digitization of public health and clinical research is likely to grow in the years to come. The COVID-19 pandemic has already played a major role in this dynamic. Indeed, social media were not only efficient to spread information and to share

XSL•FO
RenderX

diagnostic, treatment, and even follow-up protocols [80-82] but also to develop infoveillance studies to help characterize disease distribution and behaviors critical to the early stages of an outbreak [83,84] and to recruit participants in order to collect large-scale data within a short time period [85].

Still, the use of SM features and SM data for health research induces several ethical issues and limitations. Online data, such as those from Twitter, are often considered to be public, and user consent is not provided for collecting it. Moreover, ensuring privacy protection of a data set when anyone has access to vast amounts of public information is difficult because data could be reidentified [86,87]. Safety features should be used to protect users' personal and sensitive information [20] and to protect users from dangerous or fake content posted by detractors, chatbots, or social media trolls (people who purposely provoke other SM users) [88]. These kinds of behaviors can also be oriented to researchers themselves and demotivate them. Moreover, data can represent only certain users' characteristics due to researchers' self-selection or to coverage issues of underserved populations or minority groups who are disproportionately absent online (eg, older adults). This can bias the representativeness of the sample and consequently bias the findings and prevent from any generalizability [89,90]. However, it is possible to multiply platforms (cross-platforms) or to combine with other recruitment methods to minimize such bias [91]. When recruiting and providing incentives, users might be tempted to participate multiple times. Researchers should ensure that the study allows only one response from a given IP address [92,93]. A few guidelines and frameworks have already been created to guide health researchers in using social media and prevent such issues [94-98].

Strengths and Limitations of This Scoping Review

The present work used a rigorous scoping review methodology from the manual by the Joanna Briggs Institute [16] throughout the entire process. It was guided by a previously published protocol [18]. To ensure a broad search of the literature, the search strategy included two electronic bibliographic databases and the snowball technique. There are some limitations to our scoping review process. We may not have identified all relevant articles in the published literature despite attempts to be as comprehensive as possible. We limited our review to documents written in English, which may have led to missed relevant studies. Data were abstracted by one reviewer and verified by a second reviewer because of the important number of included publications.

Conclusion and Recommendations

Our findings suggest that SM hold high potential to improve and complement existing health research studies. Indeed, some SM features can complement traditional research strategies, and the growing amounts of SM data hold great opportunities in the evolution of infoveillance and infodemiology. For researchers, SM can be an effective tool at almost every step of a study, from the development, ideation, recruitment, and crowdsourcing to the dissemination of findings. Researchers should determine which type of SM best fits their objectives, as Facebook might be better for recruitment and Twitter for data collection, in order to gain time and efficiency. Last but not least, we have observed strong heterogeneity in the approaches used. We therefore recommend taking the existing guidelines into account and carefully thinking about the different ethical issues highlighted in this work before using SM for research.

Acknowledgments

This work is supported by the Luxembourg Institute of Health.

Authors' Contributions

CB and GF conducted title, abstract, and full-text screening. CB performed data extraction and data cleaning. GF performed data verification. CB and GF drafted the manuscript. CB, SS, CD, GF, AA, and CP contributed to review and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the included studies by year, country of origin, aim, population, type of social media and disease. NR: not relevant.
[XLSX File (Microsoft Excel File), 73 KB - jmir_v23i5e25736_app1.xlsx]

Multimedia Appendix 2
Coauthorship network.
[DOCX File , 12066 KB - jmir_v23i5e25736_app2.docx]

References

1. Harvey K. Encyclopedia of Social Media and Politics. United States: SAGE; 2014.
2. The Social Life of Health Information, 2011.: Pew Research Center; 2011 May 12. URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2020-06-29]

XSL•FO
RenderX

3. Bidargaddi N, Musiat P, Makinen V, Ermes M, Schrader G, Licinio J. Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. Mol Psychiatry 2017 Feb;22(2):164-169 [FREE Full text] [doi: 10.1038/mp.2016.224] [Medline: 27922603]
4. Bardus M, El Rassi R, Chahrour M, Akl E, Raslan A, Meho L, et al. The Use of Social Media to Increase the Impact of Health Research: Systematic Review. J Med Internet Res 2020 Jul 06;22(7):e15607 [FREE Full text] [doi: 10.2196/15607] [Medline: 32628113]
5. Clement J. Global digital population as of April 2020. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2020-05-01]
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]
7. Frandsen M, Thow M, Ferguson SG. The Effectiveness Of Social Media (Facebook) Compared With More Traditional Advertising Methods for Recruiting Eligible Participants To Health Research Studies: A Randomized, Controlled Clinical Trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
8. Wilson E, Kenny A, Dickson-Swift V. Using Blogs as a Qualitative Health Research Tool. International Journal of Qualitative Methods 2015 Dec 11;14(5):160940691561804. [doi: 10.1177/1609406915618049]
9. Schwab-Reese L, Hovdestad W, Tonmyr L, Fluke J. The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. Child Abuse Negl 2018 Nov;85:187-201 [FREE Full text] [doi: 10.1016/j.chiabu.2018.01.014] [Medline: 29366596]
10. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
11. Thornton L, Batterham P, Fassnacht D, Kay-Lambkin F, Calear A, 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]
12. Taylor J, Pagliari C. Comprehensive scoping review of health research using social media data. BMJ Open 2018 Dec 14;8(12):e022931 [FREE Full text] [doi: 10.1136/bmjopen-2018-022931] [Medline: 30552260]
13. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health Researchers' Use of Social Media: Scoping Review. J Med Internet Res 2019 Nov 13;21(11):e13687 [FREE Full text] [doi: 10.2196/13687] [Medline: 31719028]
14. 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]
15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
16. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping Reviews. JBI Reviewer's manual. 2019. URL: https://wiki.jbi.global/display/MANUAL/Chapter+11%3A+Scoping+reviews [accessed 2020-06-24]
17. 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 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Bour C, Schmitz S, Ahne A, Perchoux C, Dessenne C, Fagherazzi G. Scoping review protocol on the use of social media for health research purposes. BMJ Open 2021 Feb 11;11(2):e040671 [FREE Full text] [doi: 10.1136/bmjopen-2020-040671] [Medline: 33574143]
19. Gibbons J, Malouf R, Spitzberg B, Martinez L, Appleyard B, Thompson C, et al. Twitter-based measures of neighborhood sentiment as predictors of residential population health. PLoS One 2019;14(7):e0219550 [FREE Full text] [doi: 10.1371/journal.pone.0219550] [Medline: 31295294]
20. Bender J, Cyr A, Arbuckle L, Ferris L. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
21. Salathé M, Freifeld C, Mekaru S, Tomasulo A, Brownstein J. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 01;369(5):401-404 [FREE Full text] [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
22. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. J Med Internet Res 2016 Feb 26;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
23. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
24. 11.2.2 Developing the title and question - JBI Manual for Evidence Synthesis - JBI GLOBAL WIKI. URL: https://wiki.jbi.global/display/MANUAL/11.2.2+Developing+the+title+and+question [accessed 2020-07-07]
25. CADIMA. URL: https://www.cadima.info/ [accessed 2020-07-07]

XSL•FO
RenderX

26. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018 Feb 1;7(1):8. [doi: 10.1186/s13750-018-0115-5]
27. The donut and Altmetric Attention Score.: Altmetric; 2015 Jul 9. URL: https://www.altmetric.com/about-our-data/the-donut-and-score/ [accessed 2020-06-30]
28. Didegah F, Bowman TD, Holmberg K. On the differences between citations and altmetrics: An investigation of factors driving altmetrics versus citations for finnish articles. Journal of the Association for Information Science and Technology 2018 Mar 25;69(6):832-843. [doi: 10.1002/asi.23934]
29. Babu R, Vysakh C. Citations v/s Altmetric Attention Score: A Comparison of Top 10 Highly Cited Papers in Nature. 2019. URL: https://digitalcommons.unl.edu/libphilprac/2992 [accessed 2021-01-05]
30. Ortega J. Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. Journal of Informetrics 2015 Jan;9(1):39-49. [doi: 10.1016/j.joi.2014.11.004]
31. Costas, Zahedi Z, Wouters P. Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. J Assn Inf Sci Tec 2014 Jul 28;66(10):2003-2019. [doi: 10.1002/asi.23309]
32. Ford K, Albritton T, Dunn T, Crawford K, Neuwirth J, Bull S. Youth Study Recruitment Using Paid Advertising on Instagram, Snapchat, and Facebook: Cross-Sectional Survey Study. JMIR Public Health Surveill 2019 Oct 9;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
33. Colditz J, Chu K, Emery S, Larkin C, James A, Welling J, et al. Toward Real-Time Infoveillance of Twitter Health Messages. Am J Public Health 2018 Aug;108(8):1009-1014. [doi: 10.2105/AJPH.2018.304497] [Medline: 29927648]
34. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep 19;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
35. Sinnenberg L, DiSilvestro C, Mancheno C, Dailey K, Tufts C, Buttenheim A, et al. Twitter as a Potential Data Source for Cardiovascular Disease Research. JAMA Cardiol 2016 Dec 01;1(9):1032-1036 [FREE Full text] [doi: 10.1001/jamacardio.2016.3029] [Medline: 27680322]
36. Eichstaedt J, Schwartz H, Kern M, Park G, Labarthe D, Merchant R, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
37. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
38. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
39. Rudra K, Sharma A, Ganguly N, Imran M. Classifying and Summarizing Information from Microblogs During Epidemics. Inf Syst Front 2018;20(5):933-948 [FREE Full text] [doi: 10.1007/s10796-018-9844-9] [Medline: 32214879]
40. Scanfeld D, Scanfeld V, Larson E. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
41. Watson B, Robinson DHZ, Harker L, Arriola KRJ. The Inclusion of African-American Study Participants in Web-Based Research Studies: Viewpoint. J Med Internet Res 2016 Jun;18(6):e168 [FREE Full text] [doi: 10.2196/jmir.5486] [Medline: 27334683]
42. Davies B, Kotter M. Lessons From Recruitment to an Internet-Based Survey for Degenerative Cervical Myelopathy: Comparison of Free and Fee-Based Methods. JMIR Res Protoc 2018 Feb 05;7(2):e18 [FREE Full text] [doi: 10.2196/resprot.6567] [Medline: 29402760]
43. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. Internet Interv 2019 Sep;17:100246 [FREE Full text] [doi: 10.1016/j.invent.2019.100246] [Medline: 31080751]
44. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
45. The rise of social media.: Our World in Data URL: https://ourworldindata.org/rise-of-social-media [accessed 2020-07-06]
46. Clement J. Most popular social networks worldwide 2020, by reach.: Statista; 2020 Jun 23. URL: https://www.statista.com/statistics/274773/global-penetration-of-selected-social-media-sites/ [accessed 2020-07-06]
47. Anderson M, Jiang J. Teens, Social Media & Technology. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-07-06]
48. Rait M, Prochaska J, Rubinstein M. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Transl Behav Med 2015 Sep;5(3):254-259. [doi: 10.1007/s13142-015-0312-5] [Medline: 26327930]
49. Schwinn T, Hopkins J, Schinke S, Liu X. Using Facebook ads with traditional paper mailings to recruit adolescent girls for a clinical trial. Addict Behav 2017 Feb;65:207-213 [FREE Full text] [doi: 10.1016/j.addbeh.2016.10.011] [Medline: 27835860]

XSL•FO
RenderX

50. Salathé M. Digital Pharmacovigilance and Disease Surveillance: Combining Traditional and Big-Data Systems for Better Public Health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
51. Leach LS, Butterworth P, Poyser C, Batterham PJ, Farrer LM. Online Recruitment: Feasibility, Cost, and Representativeness in a Study of Postpartum Women. J Med Internet Res 2017 Mar 08;19(3):e61 [FREE Full text] [doi: 10.2196/jmir.5745] [Medline: 28274906]
52. Chu J, Snider C. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
53. 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. [doi: 10.2196/resprot.7918] [Medline: 29367186]
54. 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 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
55. Arcia A. Facebook Advertisements for Inexpensive Participant Recruitment Among Women in Early Pregnancy. Health Educ Behav 2014 Jun;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]
56. Batterham P. 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 [FREE Full text] [doi: 10.1002/mpr.1421] [Medline: 24615785]
57. Garland SM, Wark JD, Tabrizi SN, Jayasinghe Y, Moore E, Fletcher A, et al. P1-S4.32 Recruiting via social networking sites for sexual health research (assessing chlamydia and HPV knowledge). Sexually Transmitted Infections 2011 Jul 10;87(Suppl 1):A173-A174. [doi: 10.1136/sextrans-2011-050108.176]
58. Pedersen ER, Helmuth ED, Marshall GN, Schell TL, PunKay M, Kurz J. Using facebook to recruit young adult veterans: online mental health research. JMIR Res Protoc 2015 Jun 01;4(2):e63 [FREE Full text] [doi: 10.2196/resprot.3996] [Medline: 26033209]
59. Ramo D, Rodriguez T, Chavez K, Sommer M, Prochaska J. 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]
60. Youn SJ, Trinh N, Shyu I, Chang T, Fava M, Kvedar J, et al. Using online social media, Facebook, in screening for major depressive disorder among college students. International Journal of Clinical and Health Psychology 2013 Jan;13(1):74-80. [doi: 10.1016/s1697-2600(13)70010-3]
61. Santillana M, Nguyen A, Dredze M, Paul M, Nsoesie E, Brownstein J. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
62. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015 Oct;80(4):878-888 [FREE Full text] [doi: 10.1111/bcp.12746] [Medline: 26271492]
63. Tricco A, Zarin W, Lillie E, Jeblee S, Warren R, Khan P, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak 2018 Jun 14;18(1):38 [FREE Full text] [doi: 10.1186/s12911-018-0621-y] [Medline: 29898743]
64. Pappa D, Stergioulas LK. Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. Int J Data Sci Anal 2019 Feb 12;8(2):113-135 [FREE Full text] [doi: 10.1007/s41060-019-00175-3]
65. Corley CD, Cook DJ, Mikler AR, Singh KP. Using Web and social media for influenza surveillance. Adv Exp Med Biol 2010;680:559-564 [FREE Full text] [doi: 10.1007/978-1-4419-5913-3_61] [Medline: 20865540]
66. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
67. Robertson C, Yee L. Avian Influenza Risk Surveillance in North America with Online Media. PLoS One 2016;11(11):e0165688 [FREE Full text] [doi: 10.1371/journal.pone.0165688] [Medline: 27880777]
68. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6:ecurrents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
69. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
70. O'Dea B, Achilles MR, Larsen ME, Batterham PJ, Calear AL, Christensen H. The rate of reply and nature of responses to suicide-related posts on Twitter. Internet Interv 2018 Sep;13:105-107 [FREE Full text] [doi: 10.1016/j.invent.2018.07.004] [Medline: 30206524]
71. Rong J, Michalska S, Subramani S, Du J, Wang H. Deep learning for pollen allergy surveillance from twitter in Australia. BMC Med Inform Decis Mak 2019 Nov 08;19(1):208 [FREE Full text] [doi: 10.1186/s12911-019-0921-x] [Medline: 31699071]

XSL•FO
RenderX

72. Conway M, Hu M, Chapman W. Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and ConsumerGenerated Data. Yearb Med Inform 2019 Aug;28(1):208-217 [FREE Full text] [doi: 10.1055/s-0039-1677918] [Medline: 31419834]
73. Allen C, Tsou M, Aslam A, Nagel A, Gawron JM. Applying GIS and Machine Learning Methods to Twitter Data for Multiscale Surveillance of Influenza. PLoS One 2016;11(7):e0157734 [FREE Full text] [doi: 10.1371/journal.pone.0157734] [Medline: 27455108]
74. Ahne A, Orchard F, Tannier X, Perchoux C, Balkau B, Pagoto S, et al. Insulin pricing and other major diabetes-related concerns in the USA: a study of 46 407 tweets between 2017 and 2019. BMJ Open Diabetes Res Care 2020 Jun 04;8(1):e001190 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001190] [Medline: 32503810]
75. Samaras L, García-Barriocanal E, Sicilia MA. Comparing Social media and Google to detect and predict severe epidemics. Sci Rep 2020 Mar 16;10(1):4747. [doi: 10.1038/s41598-020-61686-9] [Medline: 32179780]
76. Nguyen Q, Li D, Meng H, Kath S, Nsoesie E, Li F, et al. Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. JMIR Public Health Surveill 2016 Oct 17;2(2):e158 [FREE Full text] [doi: 10.2196/publichealth.5869] [Medline: 27751984]
77. Aramburu MJ, Berlanga R, Lanza I. Social Media Multidimensional Analysis for Intelligent Health Surveillance. Int J Environ Res Public Health 2020 Mar 28;17(7):2289 [FREE Full text] [doi: 10.3390/ijerph17072289] [Medline: 32231152]
78. Bravo C, Hoffman-Goetz L. Tweeting About Prostate and Testicular Cancers: What Are Individuals Saying in Their Discussions About the 2013 Movember Canada Campaign? J Cancer Educ 2016 Sep;31(3):559-566. [doi: 10.1007/s13187-015-0838-8] [Medline: 25903054]
79. Aleksina A, Akulenka S, Lublóy Á. Success factors of crowdfunding campaigns in medical research: perceptions and reality. Drug Discov Today 2019 Jul;24(7):1413-1420. [doi: 10.1016/j.drudis.2019.05.012] [Medline: 31132416]
80. Ahmad A, Murad H. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
81. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise C, Brugnoli E, Schmidt A, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
82. González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 Pandemic. Int Braz J Urol 2020 Jul;46(suppl.1):120-124 [FREE Full text] [doi: 10.1590/S1677-5538.IBJU.2020.S121] [Medline: 32550706]
83. Rovetta A, Bhagavathula A. Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags. J Med Internet Res 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]
84. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
85. Ali S, Foreman J, Capasso A, Jones A, Tozan Y, DiClemente R. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med Res Methodol 2020 May 13;20(1):116 [FREE Full text] [doi: 10.1186/s12874-020-01011-0] [Medline: 32404050]
86. Karmaker SS, Bindschadler V, Zhai C, Gunter C. NRF: A Naive Re-identification Framework. 2018 Oct Presented at: Workshop on Privacy in the Electronic Society at the Computer and Communications Security Conference (WPES@CCS), 2018; October 2018; Toronto, Canada. [doi: 10.1145/3267323.3268948]
87. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical Issues in Social Media Research for Public Health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
88. Wekerle C, Vakili N, Stewart S, Black T. The utility of Twitter as a tool for increasing reach of research on sexual violence. Child Abuse Negl 2018 Nov;85:220-228. [doi: 10.1016/j.chiabu.2018.04.019] [Medline: 29778296]
89. Hammer M. Ethical Considerations When Using Social Media for Research. Oncol Nurs Forum 2017 Jul 01;44(4):410-412 [FREE Full text] [doi: 10.1188/17.ONF.410-412] [Medline: 28632249]
90. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
91. Wisk L, Nelson E, Magane K, Weitzman E. Clinical Trial Recruitment and Retention of College Students with Type 1 Diabetes via Social Media: An Implementation Case Study. J Diabetes Sci Technol 2019 May;13(3):445-456 [FREE Full text] [doi: 10.1177/1932296819839503] [Medline: 31010315]
92. Child R, Mentes J, Pavlish C, Phillips L. Using Facebook and participant information clips to recruit emergency nurses for research. Nurse Res 2014 Jul;21(6):16-21. [doi: 10.7748/nr.21.6.16.e1246] [Medline: 25059083]
93. Hu J, Wong K, Wang Z. Recruiting migrants for health research through social network sites: an online survey among chinese migrants in australia. JMIR Res Protoc 2015 Apr 27;4(2):e46 [FREE Full text] [doi: 10.2196/resprot.3960] [Medline: 25917837]
94. Social media guidelines for researchers - Staff home, The University of York. URL: https://www.york.ac.uk/staff/research/governance/research-policies/social-media/ [accessed 2020-07-08]
95. Townsend L, Wallace C. Social Media Research: A Guide to Ethics. URL: https://www.gla.ac.uk/media/Media_487729_smxx.pdf [accessed 2020-07-01]

XSL•FO
RenderX

96. Schillinger D, Chittamuru D, Ramírez AS. From "Infodemics" to Health Promotion: A Novel Framework for the Role of Social Media in Public Health. Am J Public Health 2020 Sep;110(9):1393-1396. [doi: 10.2105/AJPH.2020.305746] [Medline: 32552021]
97. Guidelines for Research Involving Social Media.: Ryerson University URL: https://www.ryerson.ca/content/dam/research/documents/ethics/guidelines-for-research-involving-social-media.pdf [accessed 2020-07-17]
98. Guidelines for Obtaining Consent and Assent.: Ryerson University URL: https://www.ryerson.ca/content/dam/research/documents/ethics/guidelines-for-obtaining-assent-and-consent.pdf [accessed 2020-07-17]

Abbreviations

MeSH: Medical Subject Headings
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Reviews
SM: social media

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by MA Bahrami, P Batterham, Y Hirst; comments to author 30.11.20; revised version received 15.01.21; accepted 18.03.21; published 27.05.21.

Please cite as:
Bour C, Ahne A, Schmitz S, Perchoux C, Dessenne C, Fagherazzi G
The Use of Social Media for Health Research Purposes: Scoping Review
J Med Internet Res 2021;23(5):e25736
URL: https://www.jmir.org/2021/5/e25736
doi:10.2196/25736
PMID:34042593

©Charline Bour, Adrian Ahne, Susanne Schmitz, Camille Perchoux, Coralie Dessenne, Guy Fagherazzi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

Theories Informing eHealth Implementation: Systematic Review and Typology Classification

Milena Heinsch[1,2], BSW, PhD; Jessica Wyllie[3], BBus, PhD; Jamie Carlson[3], BBus, PhD; Hannah Wells[1], BSW; Campbell Tickner[1], BSW; Frances Kay-Lambkin[1], BSc, PhD

[1]Centre for Brain and Mental Health Priority Research Centre, The University of Newcastle, Callaghan, Australia

[2]School of Humanities and Social Science, The University of Newcastle, Callaghan, Australia

[3]Newcastle Business School, The University of Newcastle, Callaghan, Australia

Corresponding Author:
Milena Heinsch, BSW, PhD
Centre for Brain and Mental Health Priority Research Centre
The University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 0249217035
Email: Milena.Heinsch@newcastle.edu.au

Abstract

Background: Theory-guided approaches to implementation science have informed translation efforts and the acceptance of eHealth (digital health) interventions in clinical care. However, there is scarce evidence on which theories are best suited to addressing the inherent complexity of eHealth implementation.

Objective: The objectives of this systematic review are to identify theories that inform and explain eHealth implementation and to classify these theories using the typology by Sovacool and Hess for theories of sociotechnical change.

Methods: An electronic search was conducted in the PsycINFO, MEDLINE, Embase, CINAHL, Scopus, Sociological Source Ultimate, Web of Science, ABI/INFORM, EBSCO, and ProQuest databases in June 2019. Studies were included if they were published between 2009 and June 2019; were written in English; reported on empirical research, regardless of study or publication type; reported on one or more theories in the context of eHealth implementation; and were published in a peer-reviewed journal. A total of 2 reviewers independently assessed the titles, abstracts, and full texts. Theories identified were classified using a typology for theories of sociotechnical change, which was considered a useful tool for ordering and analyzing the diverse theoretical approaches as a basis for future theory building.

Results: Of the 13,101 potentially relevant titles, 119 studies were included. The review identified 36 theories used to explain implementation approaches in eHealth. The most commonly used approaches were the Technology Acceptance Model (TAM) (n=33) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (n=32). These theories were primarily concerned with individual and interpersonal elements of eHealth acceptance. Less common were theories that reflect the various disorderly social processes and structural dimensions of implementation, such as the normalization process theory (n=17) and the structuration theory (n=6).

Conclusions: Theories currently informing the implementation of eHealth interventions predominantly focus on predicting or explaining end-user acceptance. Theoretical perspectives that capture the dense and intricate relationships and structures required to enact sustainable change are less well represented in the eHealth literature. Given the growing acknowledgment of the inherent complexity of eHealth implementation, future research should develop and test models that recognize and reflect the multidimensional, dynamic, and relational nature of this process.

(J Med Internet Res 2021;23(5):e18500) doi:10.2196/18500

KEYWORDS

systematic review; eHealth; digital health; mHealth; mobile phone; technology; implementation; adoption; translation; theory

XSL•FO
RenderX

Introduction

In recent years, technological innovation in health care has developed exponentially, and eHealth is now widely viewed as a significant potential contributor to improved quality of care [1,2]. However, despite much policy-level and scholarly discussions of triggering a revolution in health service delivery, problems of implementation and uptake of eHealth among both patients and service providers persist [1,3,4].

Poor uptake of eHealth (a term with contested definitions [5] but, broadly speaking, "health services and information delivered or enhanced through the internet and related technologies" [6]) is often explained in terms of barriers and facilitators [1]. In a recent study, Schreiweis et al [7] identified 77 barriers and 292 facilitators in implementing eHealth services. Similarly, a systematic review by Granja et al [8] identified 27 factors that determine the success or failure of eHealth interventions. Although studies about barriers and facilitators are important, they tend to fall short of capturing the complexity of the implementation process and the multiple *interrelated* factors that determine the translation and uptake of eHealth [1,9].

Evidence suggests that theory-informed approaches to implementation science can enhance the translation and acceptance of eHealth into clinical care [1,10-18]. Theories offer explanatory frameworks and formal heuristic devices that have the potential to move beyond the basic listing of individual facilitators and barriers to implementation, to capture *the dynamic interaction between them* [1]. As Damschroder [19] notes, theory "enables knowledge to emerge out of seeming chaos," facilitating exploration of complex relationships and interdependencies between variables that unfold in diverse and changing contexts [20]. This is of paramount importance in eHealth settings [1,18], which are characterized by a complicated interplay between patients, clinicians, the health care system, and the eHealth technology.

Many theories and models have been articulated to inform and explain eHealth implementation [15]. Despite this abundance, findings from several reviews show that only a small number of select theories have been used repeatedly across multiple publications and by several authors [21-24]. For example, a recent review by Harst et al [23] of 24 studies of end-user acceptance of telemedicine found that 2 theories accounted for 20 instances of theory use: the technology acceptance model (TAM) and the unified theory of acceptance and use of technology (UTAUT). Similarly, a review on the use of theory in eHealth weight management interventions by Willmott [24] identified 18 studies referencing a theory, of which 16 mentioned either the social cognitive theory or the transtheoretical model.

Theories most commonly used in the literature tend to emphasize individual factors, such as motivation, attitudes, and behavior, rather than the broader social and environmental factors impacting implementation [21,22,25]. This is despite evidence highlighting the multilevel nature of technology implementation in health care and the importance of targeting variables at different levels [1,26]. As Glanz and Bishop [22] noted, social and environmental factors may constrain individuals' behavior even when they are highly motivated. Therefore, the authors recommend complementing individually oriented theories with theories of social, policy, or organizational change [22].

One hindrance to this is that the current eHealth implementation literature is fragmented across multiple specialty areas and disciplines, making it difficult to locate the range of theories available [27]. To improve the selection and application of theory, it is necessary to identify an array of theories, across diverse disciplines, that have the potential to inform eHealth implementation. A further issue is that many theories contain overlapping constructs but use different terms to describe them [26]. Synthesizing theories according to their similarities would facilitate their selection and application at different levels [27].

To address these issues, we conducted a systematic review and classification of eHealth implementation theories. The review aims to address the following question: "What theories exist across disciplines that have been used to inform or explain eHealth implementation?" Theories identified by our review were classified using the typology by Sovacool and Hess [28] for theories of sociotechnical change. This typology provides an accessible and useful framework for organizing and selecting diverse theoretical options that target variables at different levels. Its use also allows the identification of areas where further theoretical development is required.

Methods

Overview

This systematic review was conducted by members of the review team in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [29]. A PRISMA checklist is available in Multimedia Appendix 1 [29]. The authors adopted a flexible approach by continuing to apply the core principles of systematic review methodology but tailoring the PRISMA guidelines to the needs of this review [30]. As such, formal quality assessment was not conducted for this review, as the perceived validity or trustworthiness of the included studies did not address the overall research question, which sought to identify the existence of theories across a broad and varied body of literature.

Search Strategy

Electronic searches of PsycINFO, MEDLINE, Embase, CINAHL, Scopus, Sociological Source Ultimate, Web of Science, ABI/INOFORM, EBSCO, and ProQuest Databases were conducted by the review team in June 2019 to identify studies that applied one or more theoretical frameworks to inform eHealth implementation. For this review, implementation was defined as "the scientific study of methods to promote the systematic uptake of research findings and other EBPs (evidence-based practices) into routine practice, and, hence, to improve the quality and effectiveness of health services." [31]. These databases were chosen because they were deemed to be likely to catalog studies and disciplines relevant to the eHealth context and the specific research question. The search was limited to studies published in the last 10 years (from 2009 to June 2019) and yielded 21,704 abstracts for initial consideration. A full list of key search terms used can be found in Multimedia

XSL•FO
RenderX

Appendix 2. All records were converted into an EndNote library and reduced to 13,101 following deduplication. Papers were then title-checked for relevance to the topic and research questions and aims before further screening by 2 independent reviewers (MH and HW) in accordance with the detailed inclusion and exclusion criteria outlined below.

Eligibility Criteria

Individual studies were included in the review if they were (1) published in the last 10 years (from 2009 to June 2019), (2) published in English, (3) outputs of empirical research or theoretical papers reporting on one or more theories in the context of eHealth implementation (this included all study types and populations), or (4) published in a peer-reviewed journal. Studies were excluded if they were (1) published before 2009, (2) not written in English, (3) studies that did not report on one or more theories applied in the context of eHealth implementation, (4) gray literature not published in a peer-reviewed journal, (5) dissertations, theses, conference proceedings, or abstracts, or (6) any form of literature review. The full eligibility criteria for this review are provided in Textbox 1.

Textbox 1. Eligibility criteria for the review.

Inclusion criteria

- Publication date from 2009 (inclusive) to June 2019
- Australian and international literature in English language
- Papers reporting on one or more theories in the context of eHealth implementation (any study type and population)
- Empirical studies (both quantitative and qualitative)
- Position, discussion, or theoretical papers
- Peer-reviewed articles

Exclusion criteria

- Publication before 2009
- Literature in non-English language
- Papers not reporting on one or more theories in the context of eHealth implementation
- Gray literature or not published in a peer-reviewed journal
- Dissertations or theses or conference proceedings or abstracts
- Literature reviews (narrative, scoping, and systematic)

Identification and Selection of Studies

A total of 2 reviewers (MH and HW) independently applied the predefined inclusion and exclusion criteria to screen for relevant studies from those obtained through database searching. To ensure accuracy, record titles and abstracts were screened manually in EndNote, and documents that did not meet the selection criteria outlined above were excluded by the reviewers. Following 2 rigorous rounds of title and abstract screening, full texts of all potentially eligible studies were examined and further screened by the 2 independent reviewers (MH and HW) using the Covidence web-based software (Veritas Health Innovation Ltd), an effective tool for assisting research teams when performing systematic reviews or meta-analyses [32]. Articles that failed to meet the selection criteria were excluded and then cross-checked to ensure transparency and accuracy surrounding the reasons given for exclusion. Any conflicts in decision making during the screening phase were resolved via discussion between reviewers or, if needed, with the research coordinator (FKL) until consensus was reached.

Data Extraction and Presentation

As the standardized extraction tool in Covidence did not meet the specific needs of this review, a modified extraction form was developed and piloted by the 2 reviewers (MH and HW) with 10 included studies tabulated and refined accordingly. The modified extraction form was tailored to include characteristics relevant to the research question. The characteristics extracted by the reviewers included (1) name of theory, (2) description, (3) instances of theory use, (4) examples of theory application, and (5) theory type. Instances of theory use refer to the number of occurrences in which a theory was used. As several studies used more than one theory, the total number of theory instances exceeded the number of papers included in the review. Examples of theory application were drawn from the literature to specify how each theory informed eHealth implementation. The reviewers then determined each theory type by drawing on the typology by Sovacool and Hess [28] for theories of sociotechnical change. This typology categorizes theories according to where they tend to *center* their analysis. The term *center* is intended to convey that a theory may involve elements of multiple types but that it approximates one ideal type above all. This typology was considered a useful tool for ordering and analyzing the diverse theoretical approaches identified, as a basis for future theory building [33].

The typology includes 5 categories: agency, structure, relations, meaning, and norms. *Agency-centered* theories relate to people's individual actions, beliefs, and attitudes, and assume that these can be explained without deeper consideration of broader social and systemic elements [28,34,35]. In contrast, *structural* theories propose that people are influenced largely by external forces

XSL•FO
RenderX

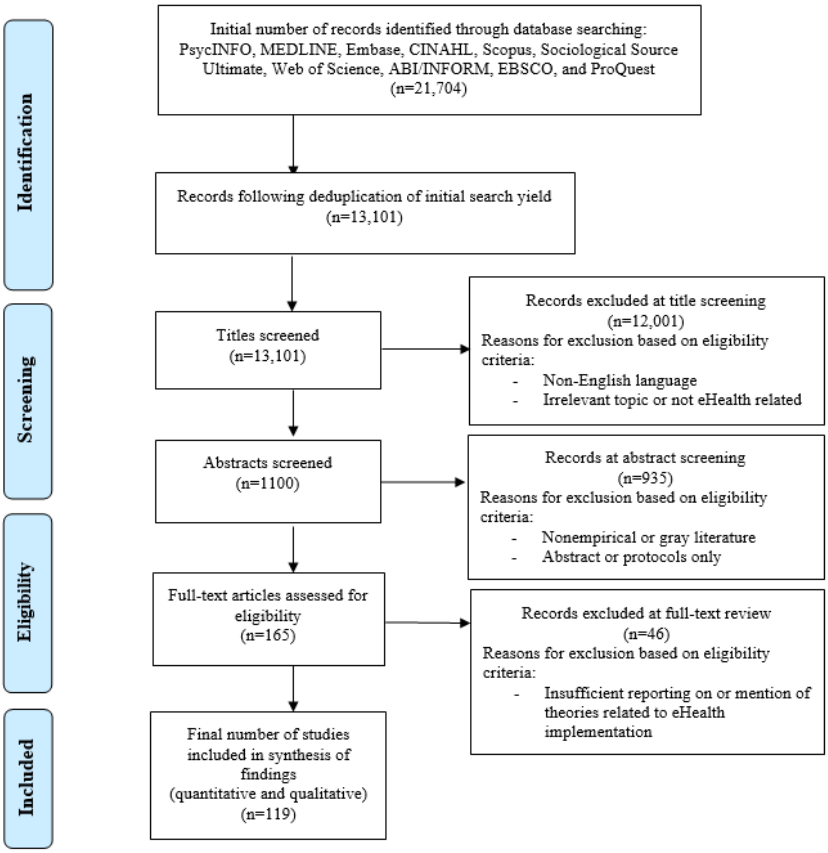
beyond their control, such as their organizational, political, or macrosocial environments [28,35]. *Relational* theories attempt to interpret the interactional processes that influence the circulation of knowledge throughout different social networks. They view technology and society as coproduced and coconstructed, with no single dimension creating change by itself [28,36]. *Meaning-centered* theories focus on language, ideas, symbolism, narratives, rhetorical visions, and other cognitive dimensions that both orient action and are changed by it. *Normative* theories offer criteria by which to assess the positive or negative impact of technology on society or on a specific group. A sixth category, *combined* theories, was added to these 5 categories. This included meta-theories that explored a combination of individual, structural, or relational frameworks. All authors (MH, JW, JC, HW, CT, and FKL) reviewed and agreed upon the classification of theories using this typology.

Results

Search Results

The electronic search of key databases resulted in 21,704 potentially eligible articles (Figure 1). This number was reduced to 13,101, following deduplication. Of these, 12,001 papers were excluded based on title screening and application of the eligibility criteria previously outlined. Key reasons for exclusion of papers at title screening included eliminating those that were in non-English language or those that reported on an irrelevant topic to the research question, for example, non-eHealth or theory-related papers. The abstracts of the remaining 1100 papers were then independently screened by reapplying the inclusion and exclusion criteria, and a further 935 papers were excluded. Key reasons for exclusion at abstract screening included nonempirical or gray literature and papers that reported abstracts or protocols only. Following a full-text review of the remaining 165 articles, an additional 46 articles were excluded because of insufficient reporting on or mention of theories related to eHealth implementation. In total, 119 articles met the full, predefined eligibility criteria and were included for data extraction and synthesis of findings. The PRISMA flowchart in Figure 1 details the process of eligibility and study selection.

Figure 1. Flowchart of studies included and excluded from the systematic review.



Theory Summary and Classification

The summary (including theory name, description, instances of theory use, and examples of application to implementation) and classification of all theories used to inform and explain eHealth implementation is provided in Table 1. In total, 36 distinct theories were identified. Classification of these theories using the typology by Sovacool and Hess [28] showed that the theories used in the literature were predominantly agency centered (19/36, 53%), followed by relational (7/36, 20%), structural (6/36, 16%), meaning (3/36, 8%), and combined theory types (1/36, 4%). No normative theories were identified.

XSL•FO
RenderX

Table 1. Summary and classification of eHealth implementation theories.

Theory	Description	Instances of theory use	Examples of theory application	Type
TAM[a]	Proposes that technology acceptance and use are affected by an individual's perceived ease of use, perceived usefulness, and subjective norms	33	• "This study tested an extended version of TAM and used this to explain the attitude of nursing staff towards using electronic patient record. The addition of external variables was shown to increase the predictive value of the model." (de Veer and Francke, 2010) [37] • "Our research introduces new variables to the TAM model in order to suit this particular study. These new variables include staff information technology experience, technical infrastructures, security concerns, and information sharing. These additional independent factors enhance the TAM's predictive power." (Zayyad and Toycan, 2018) [38]	Agency centered
UTAUT[b]	Proposes that behavioral intention to use eHealth interventions is affected by individual effort expectancy, performance expectancy, social influence, facilitating conditions, and habit	32	• "We extended the UTAUT model to investigate further context-related predictors of acceptance and postulated that eHealth literacy, which means the ability to find, evaluate, and utilize internet-based health information to health problems, and knowledge of and experience with eHealth interventions were positively related with eHealth acceptance, based on previous evidence." (Henneman et al, 2017) [39] • "To assess prospects for broad adoption of Electronic Integrated Antenatal Care, we trained midwives in the use of the system and then used the UTAUT survey to assess their intention to adopt the tool." (Markam et al, 2018) [40]	Agency centered
NPT[c]	Explains the social processes (eg, coherence, participation, collective action, and reflexive monitoring) through which eHealth interventions are operationalized	17	• "Deductive thematic and content analyses were undertaken by two independent coders [on data derived from semistructured interviews]. The NPT coding framework was used...A coding protocol was developed, trialled and refined using an additional transcript from each network." (Bagot et al, 2017) [41] • "In this paper we present a simplified set of 16 statements that express key elements of NPT, but which can be applied without a detailed knowledge of the underlying theory...we sought to better understand the ways that potential users of NPT could apply it to real world problems. Between 2006 and 2009 we engaged with multiple potential users. Engaging potential users included presentations to researchers and practitioners that linked NPT's core constructs to practical research and development problems" (May et al, 2011) [42]	Relational
DOI[d] theory	Explains how an eHealth innovation gains momentum and diffuses through a specific population. This process is affected by the innovation itself, time, channels of communication, and an individual's social system	16	• "The semistructured interview built upon Rogers' Diffusion Theory and examined the five general stages of diffusion (knowledge, persuasion, decision, implementation, and confirmation) that occurred during the clinic development. We asked respondents to describe the local mental health services before telehealth was introduced; the process by which the telehealth was introduced to—and adopted by—their organization; and the acceptance of telehealth by the community." (Brooks et al, 2012) [43] • "We used Rogers' DOI theory framework for how new innovations are adopted by organizations and Greenhalgh's subsequent work adapting the framework for health care settings. We used these frameworks to deductively explore factors that might help intervention better diffuse in each clinic setting...The analysis mapped themes identified in the qualitative data to the DOI framework described above." (Lin et al, 2016) [44]	Agency centered

XSL•FO
RenderX

Theory	Description	Instances of theory use	Examples of theory application	Type
Structuration theory	Models the relationship between agency and structure. In eHealth implementation, interventions are configured and coconstructed over time and can be adapted or revised to better accommodate different settings and needs	6	• "This study represents an empirical example of how AST[e] coupled with literature on organizational change offer a better understanding of technology implementation practices. Our findings complement past AST research, claiming that implementation success and users' change attitudes are two important outcomes associated with appropriation." (Barrett and Stephens, 2017) [45]	Relational
CFIR[f]	A meta-theoretical framework that provides an overarching typology of constructs relating to implementation, including intervention characteristics, outer setting, inner setting, and individual characteristics	6	• "The main results presented subsequently address identified barriers and facilitators influencing the implementation of internet-based patient-provider communication in 5 hospital units using CFIR to identify determinants distinguishing between high and low implementation success." (Varsi et al, 2015) [46] • "Semistructured interviews were developed based on the constructs of the CFIR which provides a pragmatic organization of theory-informed constructs known to impact implementation success across five domains...Interview transcripts were analyzed by two independent investigators using the Framework Method. This involved a largely deductive thematic analysis using a codebook based on the constructs of the CFIR." (Ware et al, 2018) [47]	Combined
ANT[g]	Posits that objects have agency and that a combination of network components (both human and nonhuman) helps create and/or influence social effects (such as implementation)	4	• "This is a qualitative study, in which ANT was used as a theoretical reference. The ANT proposes to follow and map actant' movements and the influences traversing their reciprocal connections...Our methodological reference was the 'cartography of the controversies,' considered a set of techniques to explore and visualize controversies and discussions, observing and describing social debate, especially—but not exclusively—around technical and scientific problems." (Cavalcante et al, 2019) [48]	Relational
TPB[h]	Extends the theory of reasoned action to incorporate actors' perceived control over the outcomes of their behaviors	3	• "This study integrated the Technology Acceptance Model and TPB frameworks to evaluate patient acceptance of e-health services. The Technology Acceptance Model and TPB frameworks were developed based on the theory of reasoned action." (Albar and Hoque, 2019) [49]	Agency centered
Institutional theory	Posits that an organization's environment is capable of strongly influencing the development, acceptance, and use of eHealth interventions	3	• "Due to the unique and highly institutionalized healthcare environment in the US, we therefore focus on EHR[i] adoption as an isomorphic institutional change that leads to the decision to acquire and make available electronic records for use in ambulatory services...we believe that institutional theory is a useful framework for analyzing EHR adoption." (Sherer et al, 2016 [50])	Structural
SCT[j]	Posits that the acquisition of new knowledge and perceived self-efficacy in using interventions is influenced by observing others in the context of social interactions and experiences (including the media)	3	• "The research framework encompasses three categories. The proposed bidirectional interaction between the belief types associated to these categories is based on SCT's assumption that individual behaviour is shaped by outcome expectations (behavioural factors), self-perception (personal factors) and the social and physical environment (environmental factors)." (Weeger and Gewald, 2015) [51]	Agency centered
Theory of reasoned action	A social-psychological attitude-behavior model that examines normative social influences on behavioral intention	2		Agency centered

XSL•FO
RenderX

Theory	Description	Instances of theory use	Examples of theory application	Type
			• "The theoretical background of this study focuses on the Theory of Reasoned Action...to analyze the influence of smartphones at tertiary hospitals. This study also applies the Technology Acceptance Model...Based on the accumulated knowledge from Kim and Chang, and Chang, the research model was derived..." (Moon and Chang, 2014) [52]	
IT[k]	Posits that the identity associated with a particular role is likely to drive individual decision making about eHealth interventions	2	• "We briefly reviewed the existing state of research on health IT adoption in information science and the medical informatics literatures. This is followed by a discussion of identity theories and their application to unique aspects of the healthcare context." (Mishra et al, 2012) [53]	Agency centered
PMT[l]	Explains how the perceived threat of a new eHealth intervention affects individual adoption and implementation attitudes and behaviors	2	• "To study the influence of PMT components on users' mobile health adopt intentions, we developed an integrated model based on the PMT theory and the moderating role of gender, age, and education." (Guo et al, 2015) [54]	Agency centered
Health belief model	Suggests that a personal threat, together with belief in the effectiveness of the proposed behavior, predicts the likelihood of engaging in that behavior	2	• "We applied a new adoption model that combines 3 different theories, namely, extended unified theory of acceptance and use of technology, health belief model, and the diffusion of innovation; all the 3 theories provided relevant contributions for the understanding of EHR portals. To test the research model, we used the partial least squares causal modelling approach." (Tavares and Oliveira, 2018) [55]	Agency centered
BMHSU[m]	A model for predicting and explaining factors that lead to the use of eHealth services. These can include predisposing, enabling, and needs factors	2	• "We used a multiphase, longitudinal study design. The four objectives were addressed in the four proposed stages of use. Applying Venkatesh's UTAUT and Anderson's BMHSU, we conceptualized that caregivers will go through the stages of consideration, initiation, utilization, and outcomes." (Chiu and Eysenbach, 2010) [56]	Agency centered
Sociotechnical systems theory	Proposes an interdependent relationship between the social and technical aspects of an organization (eg, human-computer interface). It suggests that these must be viewed congruently to optimize the implementation of new eHealth interventions	1	• "Drawing insight from the theoretical lens of Sociotechnical theory, the seven clusters of factors required for health-risk assessments implementation could be read as belonging to three overarching aspects: Technical (cluster 1, 2 and 3), Social-Patient (cluster 4 and 5), and Social-Provider (cluster 6 and 7)." (Ahmad et al, 2012) [57]	Relational
ECT[n]	Posits that expectations, coupled with perceived performance of an eHealth intervention, lead to post-purchase satisfaction	1	• "The conceptual framework of this research was cited from ECT. This theory was adopted extensively in the field of Marketing and Management Information System since the 1980s, though rarely applied to the healthcare field. We also used all the constructs derived from Oliver's ECT model in our research to present the original concept of this model." (Chou et al, 2012) [58]	Meaning centered
PCT[o]	Posits that past experiences of using eHealth technologies and individual assumptions in relation to the design, impact, ownership, and value of these technologies can strongly influence acceptance or reluctance toward an eHealth intervention	1	• "Arguably, PCT is a very useful framework for making more visible what lies below the surface of human problems in organisations...Consequently, this paper employs PCT as a theoretical lens to understand clinicians' reluctance to accept and use new IT systems in the NHS." (Fernando et al, 2012) [59]	Meaning centered
Resource dependence theory		1		Structural

XSL•FO
RenderX

Theory	Description	Instances of theory use	Examples of theory application	Type
	Asserts that acquiring and maintaining resources (eg, eHealth interventions) is key to organizational survival. Scarce resource availability or uncertainty about the environment motivates managers to act in ways to secure more resources and reduce their uncertainty		• "This study used the resource dependence theory to understand how the environment influences hospitals' investments in health information technology." (Tarver and Menachemi, 2018) [60]	
Theory of middle managers' role	Hypothesizes that middle managers promote implementation by fulfilling 4 roles: diffusing information, synthesizing information, mediating between strategy and day-to-day activities, and selling intervention implementation	1	• "Although the theory has received some empirical support, the extent to which it aligns with middle managers' experience in practice is unclear. The objectives of this study were to (1) assess alignment between middle managers' experience and the theory's hypothesized roles and activities and (2) elaborate on the theory with examples from middle managers' experience." (Birken et al, 2016) [61]	Structural
PAD[p] emotional state theory	Asserts that all emotional responses to physical and social environmental stimuli can be captured in 3 dimensions: pleasure (enjoyment), arousal (alertness), and dominance (control), which subsequently influence human behavior	1	• "It is the first attempt at integrating UTAUT and PAD theories to account for cognitive and affective factors in explaining technology adoption. The theory enhances the theoretical base of technical communication research by enabling theory-driven design and development of wireless health communication systems." (Alaiad and Zhou, 2017) [62]	Agency centered
Information behavior theory	Posits that information systems serve as a bridge between users and information resources. They consist of mediators (people who help users seek information and share the same social norms) and technologies (techniques and tools) that help users with the search	1	• "In this study, a critical inquiry approach was used to theorize usage behaviour through an analytic integration of three theoretical models. In our model, the driving question was as follows: What usage behaviour can be explained by Anderson's BMHSU, Venkatesh's UTAUT, and Wilson's and Chatman's information behaviour theories? We answered this question by constructing a concept map that integrates the theoretical and empirical findings. The concept map and five subthemes that influence usage and non-usage behaviour will be reported." (Chiu and Eysenbach, 2011) [63]	Agency centered
JDRM[q]	Suggests that strain is a response to imbalance between demands on individuals and the resources they have to deal with these demands	1	• "To answer our research questions, we took guidance from 2 theoretical models to ultimately derive the model...Based on the JDRM and UTAUT, we hypothesized that each of the 4 factors will positively impact provider satisfaction, and inversely relate to intention to quit." (Hysong et al, 2014) [64]	Agency centered
Cultural dimension theory	Shows the effects of a society's culture on the values of its members and how these values relate to implementation behaviors	1	• "To explore the influence of culture on e-health adoption, both TAM and Hofstede's cultural dimension model are incorporated in this study." (Hoque and Bao, 2015) [65] • "The second section [of the survey questionnaire] consists of subject perception of each variable in the model. The measurement items were adopted from prior research and modified based on the e-health context in Bangladesh." (Hoque and Bao, 2015) [65]	Meaning centered
Affect theory	Claims that there are 3 primary types of affect or emotion, including positive (joy, interest, and excitement), neutral (surprise), and negative (anger, terror, and disgust). These affective states may be advantageous or disadvantageous in users' acceptance of new information systems	1	• "Grounded in current theories of affect this study examines the role positive and negative moods play on the acceptance of a specialized telemedicine system for microbiology consultation and diagnostics, referred to as telepathology." (Djamasbi et al, 2009) [66] • "Using cognitive theories of affect, we propose an extension to TAM by arguing that users' affect plays a significant role in influencing their attitude towards a new healthcare information system." (Djamasbi et al, 2009) [67]	Agency centered

XSL•FO
RenderX

Theory	Description	Instances of theory use	Examples of theory application	Type
Activity theory	Focuses on understanding the mental capabilities of individuals by analyzing the cultural and technical aspects of human actions	1	• "We propose an integrated research model for exploring and understanding critical factors influencing physicians' intention to use computerized Clinical Practice Guidelines by incorporating activity theory (three dimensions of factors) with TAM concepts (intention as dependent variable)." (Hsiao and Chen, 2016) [68]	Agency centered
Social capital theory	Contends that social relationships are resources that can lead to the development and accumulation of human capital	1	• "This study integrated social capital theory, social cognitive theory, and TAM to develop a comprehensive behavioral model for analyzing the relationships among social capital factors (social capital theory), technological factors (TAM), and system self-efficacy (social cognitive theory) in telehealth." (Tsai, 2014) [69]	Agency centered
Contingency theory	Claims that there is no best way to organize or lead an organization or to make decisions. Instead, the optimal course of action is dependent upon the internal and external situation	1	• "In this study, we use contingency theory as a base to hypothesize how contingent factors, above and beyond traditionally considered 'dominant' factors often associated with supply-side adoption, may affect the adoption of patient portals by ambulatory-care clinics." (Baird et al, 2012) [70]	Structural
Social information processing theory	Predicts that technology-related attitudes and behaviors are not individually laden but socially constructed	1	• "When relating social information processing theory and the social influence model to organizational change situations, we see that both theories provide a framework for understanding previous scholars' arguments advocating the noteworthy role informal, coworker communication plays in effective organizational and healthcare change." (Barrett and Stephens, 2017) [71]	Relational
Consequence of modernity	Suggests that use of technologies is influenced by trust and sense of security in the absence of complete information from face-to-face interactions	1	• "Our research uses concepts from Giddens's structuration theory and consequence of modernity to understand clinical users view on telehealth service when first introduced in their work setting." (Sharma et al, 2010) [70]	Agency centered
Social worlds theory	Proposes that social worlds are self-organizing units in which people share resources, information, and assumptions about what is important and ideas about what types of activities are desirable	1	• "Using the notions of social worlds, trajectories, and boundary objects enables us to show how mobile information technology innovation in Danish home care can facilitate negotiation and collaboration across different social worlds in one setting while becoming a source of tension and conflicts in others." (Nielsen and Mengiste, 2014) [72]	Relational
Boundary objects	Boundary objects serve as interfaces between multiple social worlds and facilitate the interaction; communication; and flow of information, concepts, skills, and materials between diverse social actors	1	• "This article contributes to this emerging research domain by using notions of social worlds, trajectories, and boundary objects and applying these constructs in an empirical investigation in Danish elderly home care. Our discussion therefore focuses on two key issues: to what extent different interests among multiple social worlds have been negotiated in the trajectory of adopting and diffusing mobile IT and to what extent boundary objects have aligned the interests of stakeholders from different social worlds." (Nielsen and Mengiste, 2014) [72]	Relational
Adult learning theory	Posits that one's learning context influences learning outcomes. Thus, knowledge evolves not only through formal learning activities, such as training programs, but also through the context and culture in which they are delivered	1	• "We drew on concepts from social cognitive theory and situated cognition theory (from adult learning theory) to frame our study of training practices within the ambulatory EHR system implementation process. These theories helped us develop five propositions related to the importance of training in promoting meaningful use of EHR systems." (McAlearney et al, 2012) [73]	Agency centered
Social contagion theory		1		Agency centered

XSL•FO
RenderX

Theory	Description	Instances of theory use	Examples of theory application	Type
	A theory of collective behavior that explains how ideas and opinions spread in a social network. It holds that actors' behaviors are a function of their exposure to others' behaviors		• "Based on extensive literature review and drawing upon two theories—social contagion theory and task technology fit theory—I argue that the adoption of EHR system is contagious among health care providers; however, the contagion effect depends on the fit between the characteristics of EHR system and the characteristics of health care providers." (Gan, 2015) [74]	
TTF[r] theory	Explains how technology interacts with the tasks or activities of an organization and impacts their performance	1	• "The findings also suggest that the two factors (TTF and social contagion) are not independent and the interaction of them plays a more important role that either of them alone." (Gan, 2015) [74]	Structural
Technology organization environment theory	Predicts that technology adoption is influenced by factors relating to technological, organizational, and environmental dimensions	1	• "To investigate the factors influencing the adoption of HIS[s] in the hospitals' work processes, this study proposed the initial theoretical framework based on the combined Technology Organization Environment, institutional theory, and Human Organization Technology fit model." (Ahmadi et al, 2017) [75]	Structural

[a]TAM: technology acceptance model.
[b]UTAUT: unified theory of acceptance and use of technology.
[c]NPT: normalization process theory.
[d]DOI: diffusion of innovations.
[e]AST: adaptive structuration theory.
[f]CFIR: Consolidated Framework for Implementation Research.
[g]ANT: actor-network theory.
[h]TPB: theory of planned behavior.
[i]EHR: electronical health record.
[j]SCT: social cognitive theory.
[k]IT: identity theory.
[l]PMT: protection motivation theory.
[m]BMHSU: behavioral model of health service utilization.
[n]ECT: expectation confirmation theory.
[o]PCT: personal construct theory.
[p]PAD: pleasure, arousal and dominance.
[q]JDRM: job demands resource model.
[r]TTF: Task Technology Fit.
[s]HIS: hospital information system.

In total, 53% (19/36) of theories were classified as *agency centered*. Individual theories that occurred most frequently in the literature were the TAM by Davis and Venkatesh [76] (33 instances), UTAUT by Venkatesh [77] (32 instances), and Diffusion of Innovations Theory by Rogers [78] (16 instances). These theories were found to be primarily concerned with the individual and interpersonal elements of eHealth implementation. Although they did, to some extent, appear to consider the influence of organizational and social factors on eHealth adoption, individual attitudes, behaviors, and motivations remained the core focus of theoretical analysis. Theories classified as individual examined the adoption of eHealth either before or soon after the implementation of an intervention. However, they did not emphasize any form of user involvement in the development of an intervention. These theories tended to depict adoption as a temporally discrete and relatively immediate event, rather than as one stage in a larger multistage process. They often focused on what people were going to do soon, a decision they are about to make, or a behavior they need to alter. The diffusion of innovations theory provides an exception, as this theory considers time to be an essential factor influencing adoption [79].

A total of 20% (7/36) of theories identified in the literature were classified as *relational*. Of these, the normalization process theory (NPT) by May et al [80] occurred most frequently in the literature (17 instances), followed by structuration theory (ST) [81] (6 instances) and actor-network theory (ANT) [82,83] (4 instances). Sociotechnical systems theory, social information processing theory, social worlds theory, and boundary object theory occurred only once each in the literature. Relational theories emphasize social relations and interactions at the human-technology interface. They highlighted the complex networks of social structure and meaning in which people are embedded, proposing that the translation of knowledge is

XSL•FO
RenderX

facilitated by processes of circulation both within and across different social worlds. Some relational theories, such as ANT and ST, emphasized the role of nonhuman actors, such as computer software or programs, in transforming and mediating social relationships. These theories tended to view technology and society as coconstructed or coproduced, with no single dimension dictating change by itself. Within these theories, coproduction and implementation were often described as continuous processes, in which eHealth interventions were adapted to better accommodate different end-user settings and needs.

A total of 16% (6/36) of theories were classified as *structural*. The most common structural theory was institutional theory (IT) [84] (3 instances). Resource dependence theory, theory of middle managers' role, contingency theory, task technology fit theory, and technology organization environment theory occurred only once each in the literature. These theories conceptualized *structure* as including institutional or organizational systems as well as political, cultural, and other macrosocial environments. They often assumed that people are constrained or influenced by external forces frequently beyond their comprehension or control. For example, IT posits that organizational structures and cultural norms drive eHealth implementation, despite strong political influence.

A total of 8% (3/36) of theories were classified as *meaning centered*: expectation confirmation theory, personal construct theory, and cultural dimension theory. Each of these theories occurred only once in the literature. These theories tended to focus on the cognitive dimensions (expectations, perceptions, and beliefs) that explain people's willingness to accept the use of new health technologies. Although some meaning-centered theories, such as cultural dimension theory, have considered the influence of cultural values on the adoption and use of eHealth, these theories nonetheless centered their analysis at the individual level and were often used in combination with agency-level theories.

The Consolidated Framework for Implementation Research (CFIR) [85] was the only theory to be classified as a *combined* theory type. This theory is a meta-theoretical framework that provides a comprehensive listing of individual, social, and organizational constructs thought to influence eHealth implementation. However, it does not consider how these factors might be interrelated or how changes occur.

Discussion

Principal Findings

Evidence from a range of disciplines suggests that theory-informed approaches to implementation science are integral to the translation and implementation of eHealth into clinical care [1,10-18]. Analysis of the 119 studies included in this review identified 36 distinct theories that inform or explain eHealth implementation. However, only a few selected theories (UTAUT and TAM) were dominant, which is consistent with the findings from previous reviews [21-24]. Although these theories have been empirically proven to explain or predict certain aspects of implementation, Willmott et al [24] and Davis et al [21] caution that overreliance on *common* or *favorite* theories without direct questioning of their underlying assumptions limits progress in the field.

The typology by Sovacool and Hess [28] facilitated a closer examination of the assumptions underlying eHealth implementation theories. The findings revealed that the majority of theories were agency centered, emphasizing individual factors rather than the broader social and environmental factors impacting implementation. Although these findings were consistent with previous reviews [21,22,24,86], the wider net cast for this review provided the needed validation that this trend can be observed across multiple specialty areas and disciplines [27]. This calls into question whether theories currently being used to inform and explain the eHealth implementation adequately address the multiple and complex factors that influence the implementation process, and highlights the need for more dynamic, multilevel models of eHealth implementation [21,23,87].

This review identified a number of theories classified as relational or structural, which, to varying degrees, capture the complexity and multilevel nature of eHealth implementation. The most commonly cited relational theories were NPT, ST, and ANT. These theories recognize the important role of actors, relationships, and networks in mobilizing knowledge and embedding interventions into everyday practice. For ANT, networks are made up of both human and nonhuman *actors*, and technologies are understood to have agency and the potential to transform human interactions [88,89]. From this perspective, it may be a particularly useful theory for examining the implementation of eHealth technologies and the impacts these technologies have on human behavior. A criticism of ANT is that it has a *flat* ontology and refuses to consider institutional sources of power and inequality. Here, NPT and ST offer a possible extension, as both theories recognize the inseparable intersection between individual agents and wider social and organizational structures and norms. Structural theories also consider the influence of external forces on individual behavior and decision making. For example, IT, the most commonly used structural theory in this review, posits that an organization's environment is capable of strongly influencing the development, acceptance, and use of eHealth interventions. This theory is considered particularly relevant for application in eHealth environments, which are highly institutionalized and subject to multiple regulatory forces, high levels of professionalism, and growing network externalities that can influence adoption decisions [50].

Of particular interest was the lacuna of normative theories identified in this review. Normative theories attempt to answer whether a technology is a net positive or negative for society and individuals [28]. To do so, they often rely on evaluative criteria determined by ethics, moral studies, political ecology, or social justice. Social justice theory and sustainable development are 2 common examples of normative theories. The absence of normative theories in eHealth implementation studies is emblematic of the broader tendency of implementation science to overlook the importance of contextual factors, such as economic, social, historical, and political forces, that perpetuate inequalities in the delivery of health care services

XSL•FO
RenderX

[90]. This omission is concerning in the context of eHealth, as digital technologies have been found to exacerbate inequalities associated with older age, lower level of educational attainment, and lower socioeconomic status [91]. Future research should not shy away from normative questions of equity, justice, and sustainability and should find ways to incorporate theoretical approaches that enable exactly that.

When incorporating or combining theories, Sovacool and Hess [28] highlight the need for careful consideration of the *epistemological baggage* of different approaches. Combining multiple theoretical approaches may offer a more complete understanding or explanation, yet such combinations may mask contrasting assumptions regarding key issues [92]. For instance, are people driven primarily by their individual attitudes and motivation or do pervasive organizational cultures and social systems impose norms and values that shape people's behavior, making individual characteristics relatively unimportant? These challenges may account for the tendency of theories to target variables at the same level. One exception was the CFIR framework, which was the sole theory that provided a *menu* of constructs at different levels for researchers to choose from. However, although CFIR recognizes the multilevel nature of eHealth implementation, it does not consider the relationship between constructs or how change takes place, leading Nilsen [92] to contend that it should not be considered a theory at all. Further research is needed to explore how diverse theoretical perspectives can be brought together in ways that capture the dynamic interaction between constructs [1], while avoiding disconnects and incompatibilities [28].

Limitations

This study has several limitations. First, papers not published in English were excluded, which may indicate a selection bias. The decision to keep the research question and inclusion criteria for this review broad resulted in a high yield of papers and, to some extent, reduced the specificity of search results. This decision was made to ensure the identification of the full spectrum of theories being used to inform and explain eHealth implementation. Restriction of inclusion criteria in previous systematic reviews [24] led to the omission of a number of key theories that provide a more comprehensive explanation of the various constituents of the implementation processes. A further limitation is that the protocol for this systematic review was not registered. However, every care was taken to ensure compliance with the core principles of the systematic review methodology. As Mallett [30] noted, systematic reviews do not constitute a homogenous approach, and researchers may adopt a more flexible approach that better suits their research purpose while continuing to comply with the principles for conducting a systematic review. Finally, the literature search for this review was conducted in June 2019. Given the rapid rate of publication in the field of eHealth, it is likely that recent relevant articles have not been included. As completing an updated search was not feasible for the research team, we suggest that future studies must continue to identify theories used to inform and explain the implementation of eHealth interventions.

Conclusions

This systematic review identified 36 theories that are being used to inform and explain eHealth implementation and classified these theories using the categories adapted from the typology by Sovacool and Hess [28] for theories of sociotechnical change. The results highlight the dominance of theories that focus mainly on individual readiness to accept health technologies rather than the various disorderly social processes or systemic dimensions of implementation. This calls into question whether theories currently being used to inform and explain eHealth implementation adequately address the multiple and multilevel factors that influence the implementation process. Nonetheless, this review identified a number of theories classified as relational, structural, or combined, which, to varying degrees, capture the complex interactions within a wider organization and policy system. Although less prominent in the literature, these theories may be particularly applicable to the implementation of eHealth in health settings and services.

Acknowledgments

This systematic review is part of the eCliPSE (Electronic Clinical Pathways to Service Excellence) project, which is funded by a National Health and Medical Research Council partnership project grant (APP117181) and Beyond Blue.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 80 KB - jmir_v23i5e18500_app1.doc]

Multimedia Appendix 2
Key search terms used.
[DOCX File , 16 KB - jmir_v23i5e18500_app2.docx]

References

XSL•FO
RenderX

1. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
2. Ossebaard HC, Van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
3. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
4. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015;1:1-10 [FREE Full text] [doi: 10.1177/2055207615595335] [Medline: 29942544]
5. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
6. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
8. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
9. Heinsch M. Exploring the potential of interaction models of research use in social work. Br J Soc Work 2018;48(2):468-486. [doi: 10.1093/bjsw/bcx034]
10. Jennings HM, Morrison J, Akter K, Kuddus A, Ahmed N, Shaha SK, et al. Developing a theory-driven contextually relevant mHealth intervention. Glob Health Action 2019;12(1):1550736 [FREE Full text] [doi: 10.1080/16549716.2018.1550736] [Medline: 31154988]
11. 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]
12. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
13. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10(53):1-13 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
14. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]
15. Kitson A. Moving on…. J Clin Nurs 2018 Jun 10;27(11-12):2175-2176. [doi: 10.1111/jocn.14383] [Medline: 29633468]
16. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
17. Ridde V. Need for more and better implementation science in global health. BMJ Glob Health 2016;1(2):e000115 [FREE Full text] [doi: 10.1136/bmjgh-2016-000115] [Medline: 28588947]
18. Van Belle S, van de Pas R, Marchal B. Towards an agenda for implementation science in global health: there is nothing more practical than good (social science) theories. BMJ Glob Health 2017;2(2):e000181 [FREE Full text] [doi: 10.1136/bmjgh-2016-000181] [Medline: 28608872]
19. Damschroder LJ. Clarity out of chaos: use of theory in implementation research. Psychiatry Res 2020 Jan;283:112461 [FREE Full text] [doi: 10.1016/j.psychres.2019.06.036] [Medline: 31257020]
20. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Dec 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
21. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015 Sep;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
22. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
23. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
24. Willmott T, Pang B, Rundle-Thiele S, Badejo A. Reported theory use in electronic health weight management interventions targeting young adults: a systematic review. Health Psychol Rev 2019 Sep;13(3):295-317. [doi: 10.1080/17437199.2019.1625280] [Medline: 31161877]
25. Brennan L, Previte J, Fry M. Social marketing's consumer myopia. J Soc Mark 2016 Jul 11;6(3):219-239. [doi: 10.1108/jsocm-12-2015-0079]

XSL•FO
RenderX

26. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
27. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
28. Sovacool BK, Hess DJ. Ordering theories: typologies and conceptual frameworks for sociotechnical change. Soc Stud Sci 2017 Oct 23;47(5):703-750 [FREE Full text] [doi: 10.1177/0306312717709363] [Medline: 28641502]
29. Moher D, Liberati A, Tetzlaff J, Altman DG. 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]
30. Mallett R, Hagen-Zanker J, Slater R, Duvendack M. The benefits and challenges of using systematic reviews in international development research. J Dev Effect 2012 Sep;4(3):445-455. [doi: 10.1080/19439342.2012.711342]
31. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015 Sep 16;3:32 [FREE Full text] [doi: 10.1186/s40359-015-0089-9] [Medline: 26376626]
32. Helfer B, Samara M, Leucht S. Experiences with covidence in preparing a comprehensive systematic review. In: Filtering the Information Overload for Better Decisions. 2015 Presented at: 23rd Cochrane Colloquium; Oct 3-7, 2015; Vienna, Austria.
33. O'Raghallaigh P, Sammon D, Murphy C. Map-making and theory-building: principles for achieving presentation and content affectiveness. In: Proceedings of the International Conference on Information Systems, ICIS 2010. 2010 Presented at: International Conference on Information Systems, ICIS 2010; December 12-15, 2010; Saint Louis, Missouri, USA,.
34. Hess DJ. Neoliberalism and the history of STS theory: toward a reflexive sociology. Soc Epistemol 2013 Apr;27(2):177-193. [doi: 10.1080/02691728.2013.793754]
35. Jackson T. Motivating sustainable consumption: a review of evidence on consumer behaviour and behavioural change. The Sustainable Development Research Network, Surrey. 2005. URL: https://timjackson.org.uk/wp-content/uploads/2018/04/Jackson.-2005.-Motivating-Sustainable-Consumption.pdf [accessed 2019-06-30]
36. Rutherford J, Coutard O. Urban energy transitions: places, processes and politics of socio-technical change. Urban Stud 2014 Apr 22;51(7):1353-1377. [doi: 10.1177/0042098013500090]
37. de Veer AJ, Francke AL. Attitudes of nursing staff towards electronic patient records: a questionnaire survey. Int J Nurs Stud 2010 Jul;47(7):846-854. [doi: 10.1016/j.ijnurstu.2009.11.016] [Medline: 20022007]
38. Zayyad MA, Toycan M. Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of Nigerian hospitals from the perspective of healthcare professionals. PeerJ 2018;6:1-15 [FREE Full text] [doi: 10.7717/peerj.4436] [Medline: 29507830]
39. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
40. Markam H, Hochheiser H, Kuntoro K, Notobroto HB. Exploring midwives' need and intention to adopt electronic integrated antenatal care. Perspect Health Inf Manag 2018(Winter):1-21 [FREE Full text] [Medline: 29618961]
41. Bagot KL, Cadilhac DA, Bladin CF, Watkins CL, Vu M, Donnan GA, VSTASTUTE investigators. Integrating acute stroke telemedicine consultations into specialists' usual practice: a qualitative analysis comparing the experience of Australia and the United Kingdom. BMC Health Serv Res 2017 Nov 21;17(1):1-8 [FREE Full text] [doi: 10.1186/s12913-017-2694-1] [Medline: 29157233]
42. May CR, Finch T, Ballini L, MacFarlane A, Mair F, Murray E, et al. Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. BMC Health Serv Res 2011 Sep 30;11:245 [FREE Full text] [doi: 10.1186/1472-6963-11-245] [Medline: 21961827]
43. Brooks E, Manson SM, Bair B, Dailey N, Shore JH. The diffusion of telehealth in rural American Indian communities: a retrospective survey of key stakeholders. Telemed J E Health 2012;18(1):60-66 [FREE Full text] [doi: 10.1089/tmj.2011.0076] [Medline: 22082106]
44. Lin CP, Guirguis-Blake J, Keppel GA, Dobie S, Osborn J, Cole AM, et al. Using the diffusion of innovations theory to assess socio-technical factors in planning the implementation of an electronic health record alert across multiple primary care clinics. J Innov Health Inform 2016 Apr 15;23(1):450-458 [FREE Full text] [doi: 10.14236/jhi.v23i1.157] [Medline: 27348488]
45. Barrett AK, Stephens KK. The pivotal role of change appropriation in the implementation of health care technology. Manag Commun Q 2016 Dec 18;31(2):163-193. [doi: 10.1177/0893318916682872]
46. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262. [doi: 10.2196/jmir.5091] [Medline: 26582138]

XSL•FO
RenderX

47. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the implementation of a mobile phone-based telemonitoring program: longitudinal study guided by the consolidated framework for implementation research. JMIR Mhealth Uhealth 2018 Jul 31;6(7):13 [FREE Full text] [doi: 10.2196/10768] [Medline: 30064970]
48. Cavalcante RB, Esteves CJ, Gontijo TL, Brito MJ, Guimarães EA, Barbosa SP. Computerization of primary health care in Brazil: the network of actors. Rev Bras Enferm 2019;72(2):337-344 [FREE Full text] [doi: 10.1590/0034-7167-2018-0381] [Medline: 31017194]
49. AlBar AM, Hoque MR. Patient acceptance of e-health services in Saudi-Arabia: an integrative perspective. Telemed J E Health 2019 Sep;25(9):847-852. [doi: 10.1089/tmj.2018.0107] [Medline: 30452333]
50. Sherer SA, Meyerhoefer CD, Peng L. Applying institutional theory to the adoption of electronic health records in the U.S. Inf Manag 2016 Jul;53(5):570-580. [doi: 10.1016/j.im.2016.01.002]
51. Weeger A, Gewald H. Acceptance and use of electronic medical records: an exploratory study of hospital physicians' salient beliefs about HIT systems. Health Syst 2015;4(1):64-81. [doi: 10.1057/hs.2014.11]
52. Moon BC, Chang H. Technology acceptance and adoption of innovative smartphone uses among hospital employees. Healthc Inform Res 2014 Oct;20(4):304-312 [FREE Full text] [doi: 10.4258/hir.2014.20.4.304] [Medline: 25405067]
53. Mishra AN, Anderson C, Angst CM, Agarwal R. Electronic health records assimilation and physician identity evolution: an identity theory perspective. Inf Syst Res 2012 Sep;23(3-part-1):738-760. [doi: 10.1287/isre.1110.0407]
54. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating m-Health acceptance from a protection motivation theory perspective: gender and age differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]
55. Tavares J, Oliveira T. New integrated model approach to understand the factors that drive electronic health record portal adoption: Cross-Sectional National Survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
56. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010 Nov 23;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
57. Ahmad F, Norman C, O'Campo P. What is needed to implement a computer-assisted health risk assessment tool? An exploratory concept mapping study. BMC Med Inform Decis Mak 2012 Dec 19;12:149 [FREE Full text] [doi: 10.1186/1472-6947-12-149] [Medline: 23253913]
58. Chou HK, Lin IC, Woung LC, Tsai MT. Engagement in e-learning opportunities: an empirical study on patient education using expectation confirmation theory. J Med Syst 2012 Jun;36(3):1697-1706. [doi: 10.1007/s10916-010-9630-9] [Medline: 21104303]
59. Fernando S, Choudrie J, Lycett M, de Cesare S. Hidden assumptions and their influence on clinicians' acceptance of new IT systems in the NHS. Inf Syst Front 2012;14(2):279-299. [doi: 10.1007/s10796-010-9238-0]
60. Tarver WL, Menachemi N. Environmental market factors associated with electronic health record adoption among cancer hospitals. Health Care Manage Rev 2018;43(4):303-314. [doi: 10.1097/HMR.0000000000000149] [Medline: 28225447]
61. Birken SA, DiMartino LD, Kirk MA, Lee SY, McClelland M, Albert NM. Elaborating on theory with middle managers' experience implementing healthcare innovations in practice. Implement Sci 2016;11(2):1-15 [FREE Full text] [doi: 10.1186/s13012-015-0362-6] [Medline: 26729367]
62. Alaiad A, Zhou L. Patients' adoption of WSN-based smart home healthcare systems: an integrated model of facilitators and barriers. IEEE Trans Profess Commun 2017 Mar;60(1):4-23. [doi: 10.1109/TPC.2016.2632822]
63. Chiu TM, Eysenbach G. Theorizing the health service usage behavior of family caregivers: a qualitative study of an internet-based intervention. Int J Med Inform 2011 Nov;80(11):754-764. [doi: 10.1016/j.ijmedinf.2011.08.010] [Medline: 21958550]
64. Hysong SJ, Spitzmuller C, Espadas D, Sittig DF, Singh H. Electronic alerts and clinician turnover: the influence of user acceptance. Am J Manag Care 2014 Nov;20(11 Spec No. 17):520-530 [FREE Full text] [Medline: 25811826]
65. Hoque MR, Bao Y. Cultural influence on adoption and use of e-Health: evidence in Bangladesh. Telemed J E Health 2015 Oct;21(10):845-851. [doi: 10.1089/tmj.2014.0128] [Medline: 26348844]
66. Djamasbi S, Fruhling AL, Loiacono ET. The influence of affect, attitude and usefulness in the acceptance of telemedicine systems. Information Systems and Quantitative Analysis Faculty Publications. 2009. URL: https://core.ac.uk/download/pdf/232755549.pdf [accessed 2021-05-19]
67. Tsai CH. Integrating social capital theory, social cognitive theory, and the technology acceptance model to explore a behavioral model of telehealth systems. Int J Environ Res Public Health 2014 May 07;11(5):4905-4925 [FREE Full text] [doi: 10.3390/ijerph110504905] [Medline: 24810577]
68. Hsiao JL, Chen RF. Critical factors influencing physicians' intention to use computerized clinical practice guidelines: an integrative model of activity theory and the technology acceptance model. BMC Med Inform Decis Mak 2016;16:15 [FREE Full text] [doi: 10.1186/s12911-016-0241-3] [Medline: 26772169]
69. Baird A, Furukawa MF, Raghu TS. Understanding contingencies associated with the early adoption of customer-facing web portals. J Manag Inf Syst 2014 Dec 08;29(2):293-324. [doi: 10.2753/MIS0742-1222290210]
70. Sharma U, Barnett J, Clarke M. Stud Health Technol Inform 2010;160(Pt 1):545-549. [Medline: 20841746]

XSL•FO
RenderX

71. Barrett AK, Stephens KK. Making Electronic Health Records (EHRs) work: informal talk and workarounds in healthcare organizations. Health Commun 2017 Aug;32(8):1004-1013. [doi: 10.1080/10410236.2016.1196422] [Medline: 27463257]
72. Nielsen JA, Mengiste SA. Analysing the diffusion and adoption of mobile IT across social worlds. Health Informatics J 2014 Jun;20(2):87-103 [FREE Full text] [doi: 10.1177/1460458213481688] [Medline: 24810724]
73. McAlearney AS, Robbins J, Kowalczyk N, Chisolm DJ, Song PH. The role of cognitive and learning theories in supporting successful EHR system implementation training: a qualitative study. Med Care Res Rev 2012 Jun;69(3):294-315. [doi: 10.1177/1077558711436348] [Medline: 22451617]
74. Gan Q. Is the adoption of electronic health record system “contagious”? Health Policy Technol 2015 Jun;4(2):107-112. [doi: 10.1016/j.hlpt.2015.02.009]
75. Ahmadi H, Nilashi M, Shahmoradi L, Ibrahim O. Hospital Information System adoption: expert perspectives on an adoption framework for Malaysian public hospitals. Comput Hum Beh 2017 Feb;67:161-189. [doi: 10.1016/j.chb.2016.10.023]
76. Davis F, Venkatesh V. Toward preprototype user acceptance testing of new information systems: implications for software project management. IEEE Trans Eng Manage 2004 Feb;51(1):31-46. [doi: 10.1109/TEM.2003.822468]
77. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf Syst Res 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
78. Rogers E. Diffusion of Innovations. New York: Free Press; 2003:1-576.
79. Anttila M, Välimäki M, Koivunen M, Luukkaala T, Kaila M, Pitkänen A, et al. Adoption of an internet-based patient education programme in psychiatric hospitals. J Psychiatr Ment Health Nurs 2011 Dec;18(10):914-923. [doi: 10.1111/j.1365-2850.2011.01765.x] [Medline: 22070578]
80. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
81. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Oakland, California: University of California Press; 1986:1-417.
82. Latour B. Science in Action – How to Follow Scientists & Engineers Through Society. Cambridge, Massachusetts, United States: Harvard University Press; 1988:1-282.
83. Latour B, Woolgar S, Salk J. Laboratory Life: The Construction of Scientific Facts. Princeton, NJ: Princeton University Press; 1986:1-296.
84. Meyer JW, Rowan B. Institutionalized organizations: formal structure as myth and ceremony. Am J Sociol 1977 Sep;83(2):340-363. [doi: 10.1086/226550]
85. 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;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
86. McIsaac JD, Penney TL, Storey KE, Sigfridson L, Cunningham J, Kuhle S, et al. Integrated knowledge translation in population health intervention research: a case study of implementation and outcomes from a school-based project. Health Res Policy Syst 2018 Aug 02;16(1):72 [FREE Full text] [doi: 10.1186/s12961-018-0351-8] [Medline: 30068359]
87. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
88. Callon M. The sociology of an actor-network: the case of the electric vehicle. In: Callon M, Law J, Rip A, editors. Mapping the Dynamics of Science and Technology. London: Palgrave Macmillan; 1986:19-34.
89. Latour B. Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker W, Law J, editors. Shaping Technology-Building Society: Studies in Sociotechnical Change. Cambridge, MA: The MIT Press; 1992:225-259.
90. Baumann AA, Cabassa LJ. Reframing implementation science to address inequities in healthcare delivery. BMC Health Serv Res 2020 Mar 12;20(1):190 [FREE Full text] [doi: 10.1186/s12913-020-4975-3] [Medline: 32164706]
91. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019 Oct 01;29(3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
92. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]

Abbreviations

ANT: actor-network theory
CFIR: Consolidated Framework for Implementation Research
eCliPSE: Electronic Clinical Pathways to Service Excellence
IT: institutional theory
NPT: normalization process theory
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

ST: structuration theory
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach, Q Zeng; submitted 01.03.20; peer-reviewed by DA Hepworth, P Chow, S Nepal, CJ(Wu, T Willmott; comments to author 28.08.20; revised version received 22.10.20; accepted 15.04.21; published 31.05.21.

Please cite as:
Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F
Theories Informing eHealth Implementation: Systematic Review and Typology Classification
J Med Internet Res 2021;23(5):e18500
URL: https://www.jmir.org/2021/5/e18500
doi:10.2196/18500
PMID:34057427

©Milena Heinsch, Jessica Wyllie, Jamie Carlson, Hannah Wells, Campbell Tickner, Frances Kay-Lambkin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2021. 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

COVID-19, Social Isolation, and Mental Health Among Older Adults: A Digital Catch-22

Theodore D Cosco[1,2], PhD; Karen Fortuna[3], PhD; Andrew Wister[1], PhD; Indira Riadi[1], BSc; Kevin Wagner[1], BA; Andrew Sixsmith[4], PhD

[1]Gerontology Research Center, Simon Fraser University, Vancouver, BC, Canada

[2]Oxford Institute of Population Ageing, University of Oxford, Oxford, United Kingdom

[3]Dartmouth College, Hanover, NH, United States

[4]STAR Institute, Simon Fraser University, Vancouver, BC, Canada

Corresponding Author:
Theodore D Cosco, PhD
Gerontology Research Center
Simon Fraser University
2800 Harbour Centre, 515 West Hastings Street
Vancouver, BC,
Canada
Phone: 1 778 782 5065
Email: tdcosco@cantab.net

Abstract

One of the most at-risk groups during the COVID-19 crisis is older adults, especially those who live in congregate living settings and seniors' care facilities, are immune-compromised, and/or have other underlying illnesses. Measures undertaken to contain the spread of the virus are far-reaching, and older adults were among the first groups to experience restrictions on face-to-face contact. Although reducing viral transmission is critical, physical distancing is associated with negative psychosocial implications, such as increased rates of depression and anxiety. Promising evidence suggests that participatory digital co-design, defined as the combination of user-centered design and community engagement models, is associated with increased levels of engagement with mobile technologies among individuals with mental health conditions. The COVID-19 pandemic has highlighted shortcomings of existing technologies and challenges in their uptake and usage; however, strategies such as co-design may be leveraged to address these challenges both in the adaptation of existing technologies and the development of new technologies. By incorporating these strategies, it is hoped that we can offset some of the negative mental health implications for older adults in the context of physical distancing both during and beyond the current pandemic.

(J Med Internet Res 2021;23(5):e21864) doi:10.2196/21864

KEYWORDS

social isolation; mental health; COVID-19; technology; older adult; psychology; digital health

Older adults constitute one of the most at-risk groups during the ongoing COVID-19 crisis, especially those who live in congregate living settings and seniors' care facilities, are immune-compromised, and/or have other underlying illnesses [1]. Measures undertaken to contain the spread of the virus are far-reaching, and older adults were among the first groups to have experienced restrictions with regard to face-to-face contact. Although reducing viral transmission is critical, physical distancing is known to have significant negative psychosocial implications, such as increased rates of depression and anxiety, particularly for older adults [2].

Apart from the COVID-19 outbreak, older adults, in general, experience the highest rates of social isolation and loneliness, which are associated with a range of negative mental health outcomes [3]. Physical distancing measures may have negative impacts on older adults living in the community, especially for those who rely on formal or informal support and care from family and/or friends, those who utilize religious or community centers as social hubs, as well as those who live in supportive or assisted living centers (where gathering in common spaces used for meals and other activities must cease as a result of physical distancing measures). This double burden of age-related diminishing social spheres and the implementation of measures that require physical distancing and enforced isolation presents significant mental health risks for older people.

XSL•FO
RenderX

In response to "stay at home" orders, sweeping transitions to web-based communication platforms have been adopted as a means of fostering and maintaining social connectedness; however, older adults use these technologies at disproportionately lower rates than younger individuals, both during and beyond the COVID-19 pandemic. This digital divide, which involves barriers to digital technologies [4], is a major inhibitive force in facilitating digital social connectedness among older adults while following the stay-at-home orders. The first level of the digital divide includes barriers to uptake and access to technology (eg, financial limitations), whereas the second level concerns barriers to the usage of these technologies (eg, functional impairments in manual dexterity) [5]. Most videoconferencing or communication platforms have not been developed with the older adult in mind, thereby leading to low rates of uptake before the COVID-19 pandemic and difficulty in accessing and using these platforms during the pandemic.

Consequently, we find ourselves in a catch-22 situation where negative mental health implications of social isolation and loneliness among older adults could potentially be mitigated by the use of digital solutions, but only if the person already has the technological knowledge, desire, and access to use these technologies. Furthermore, without the capacity to interact face-to-face with older adults during the COVID-19 outbreak, it can be challenging, albeit possible, to provide guidance on how to use technologies, an issue further compounded by physiological challenges (eg, reduced visual acuity, manual dexterity, and cognitive impairment). Despite its huge potential, the impact of implementing technologies to reduce social isolation and loneliness has been limited in terms of real-world products and services [6]. Unfortunately, it is often the individuals who would benefit most from digital solutions who are the most vulnerable and least likely to access these technologies. Moving forward, there is a need for a major shift in research culture away from a researcher- and technology-driven agenda to one that focuses on real-world problems and solutions [6,7]. The digital divide spans a considerable breadth of issues within the context of the pandemic; therefore, we will focus on the challenge of technology development that is not inclusive of older adults' perspectives.

In the short term, we must aim to foster the uptake of existing technologies through trainings designed for older adults and address normal age-related changes that impede technology use (ie, changes to memory, eyesight, or mobility), with longer-term objectives to incorporate older adults throughout the software development lifecycle of new technologies. Within the context of the current challenges presented by physical distancing, repurposing or reframing existing technologies may be effective in mitigating social isolation and feelings of loneliness among the older population. For example, the telephone may not be the leading edge of innovation, but a landline phone call is a demonstrated and effective means of communication and human connection [8]. With respect to new technologies—that may be critical in addressing future pandemics—incorporating digital peer support is one of the best ways to foster technology engagement [9]. Through these live or automated peer support services that can be delivered through various digital platforms, such as peer-to-peer networks on social media or peer-delivered interventions supported with smartphone apps, older adults can develop the tools needed to actively engage with technology. The similarity, bond, and trust within this relationship supports reciprocal accountability and engagement, which may promote engagement as older adults model and teach others how to use technology. For example, older adult peer support specialists with lived experience of aging with a mental health condition (and commonly a physical health condition), who are trained and accredited by their respective state to provide Medicaid and/or Medicare reimbursable telehealth services, can use their personal experiences to help address challenges of normal aging and technology use (eg, changes to eyesight, fine motor skills, and hearing). Through such support from individuals with lived experiences of aging, it is hoped that greater technology use and uptake can be fostered.

Promising evidence suggests that participatory digital co-design, defined as the combination of user-centered design and community engagement models, shows high levels of engagement with mobile technologies among individuals with mental health conditions [9]. The COVID-19 pandemic has certainly highlighted the shortcomings of existing technologies and challenges in their uptake and usage [10]; however, strategies such as co-design may be leveraged to address these challenges both in the adaptation of existing technologies and the development of new technologies.

In populations where individuals are not living with mental illness, participatory digital co-design has led to promising developments with the potential for more relevant research with wider impact, better internal validity, and rapid translation of research into practice—and greater engagement [9]; therefore, it has the potential to reduce the digital divide [11]. By incorporating the aforementioned strategies, it is hoped that we can offset some of the negative mental health implications for older adults in the context of physical distancing, both during and beyond the COVID-19 pandemic.

Acknowledgments

TDC is a Michael Smith Foundation for Health Research Scholar (SCH-2020-0490). KF is funded by a NIMH K01 award (K01MH117496).

Conflicts of Interest

KF offers consulting services through Social Wellness. She is also affiliated with Trusst Health, Inc and InquisitHealth.

XSL•FO
RenderX

References

1. 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]
2. Robb CE, de Jager CA, Ahmadi-Abhari S, Giannakopoulou P, Udeh-Momoh C, McKeand J, et al. Associations of social isolation with anxiety and depression during the early COVID-19 pandemic: A survey of older adults in London, UK. Front Psychiatry 2020;11:591120 [FREE Full text] [doi: 10.3389/fpsyt.2020.591120] [Medline: 33132942]
3. Coyle CE, Dugan E. Social isolation, loneliness and health among older adults. J Aging Health 2012 Dec;24(8):1346-1363. [doi: 10.1177/0898264312460275] [Medline: 23006425]
4. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015 Mar;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]
5. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
6. Sixsmith A, Mihailidis A, Simeonov D. Aging and technology: Taking the research into the real world. Public Policy & Aging Report 2017 Jun;27(2):74-78. [doi: 10.1093/ppar/prx007]
7. Cosco TD, Firth J, Vahia I, Sixsmith A, Torous J. Mobilizing mHealth data collection in older adults: Challenges and opportunities. JMIR Aging 2019 Mar 19;2(1):e10019 [FREE Full text] [doi: 10.2196/10019] [Medline: 31518253]
8. Van Orden KA, Bower E, Lutz J, Silva C, Gallegos AM, Podgorski CA, et al. Strategies to promote social connections among older adults during 'social distancing' restrictions. Am J Geriatr Psychiatry. Epub ahead of print posted online on May 18, 2020 2020 [FREE Full text] [doi: 10.1016/j.jagp.2020.05.004] [Medline: 32425473]
9. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: Systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
10. Moore RC, Hancock JT. Older Adults, Social Technologies, and the Coronavirus Pandemic: Challenges, Strengths, and Strategies for Support. Social Media + Society 2020 Aug 11;6(3). [doi: 10.1177/2056305120948162]
11. 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]

Edited by C Basch; submitted 26.06.20; peer-reviewed by M O'Connell, M Feijt, C Son; comments to author 17.11.20; revised version received 08.12.20; accepted 15.04.21; published 06.05.21.

Please cite as:
Cosco TD, Fortuna K, Wister A, Riadi I, Wagner K, Sixsmith A
COVID-19, Social Isolation, and Mental Health Among Older Adults: A Digital Catch-22
J Med Internet Res 2021;23(5):e21864
URL: https://www.jmir.org/2021/5/e21864
doi:10.2196/21864
PMID:33891557

©Theodore D Cosco, Karen Fortuna, Andrew Wister, Indira Riadi, Kevin Wagner, Andrew Sixsmith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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

Enhancing Data Linkage to Break the Chain of COVID-19 Spread: The Taiwan Experience

Wei-Cheng Lo[1*], PhD; Fu-Chung Wang[2,3*], PhD; Li-Yin Lin[1,4,5], PhD; Hong-Wei Jyan[6], MSc; Hsuan-Chien Wu[7], BA; Yao-Liang Huang[8], MSc; I-Ming Parng[2], MSc; Hung-Yi Chiou[1,4,5], PhD

[1]Master Program in Applied Epidemiology, College of Public Health, Taipei Medical University, Taipei, Taiwan
[2]Department of Information Management, Ministry of Health and Welfare, Taipei, Taiwan
[3]College of Health Technology, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[4]Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, Taiwan
[5]School of Public Health, College of Public Health, Taipei Medical University, Taipei, Taiwan
[6]Department of Cyber Security, Executive Yuan, Taipei, Taiwan
[7]Centers for Disease Control, Ministry of Health and Welfare, Taipei, Taiwan
[8]National Immigration Agency, Ministry of the Interior, Taipei, Taiwan
[*]these authors contributed equally

Corresponding Author:
Hung-Yi Chiou, PhD
Institute of Population Health Sciences
National Health Research Institutes
No 35 Keyan Road
Zhunan, Miaoli County, 350
Taiwan
Phone: 1 886 2 2736 1661 ext 6512
Email: hychiou@tmu.edu.tw

Abstract

Digital technology has been widely used in health care systems and disease management, as well as in controlling the spread of COVID-19. As one of the most successful countries in combating the COVID-19 pandemic, Taiwan has successfully used digital technology to strengthen its efforts in controlling the COVID-19 pandemic. Taiwan has a well-established National Health Insurance System (NHIS), which provides a great opportunity to develop a nationwide data linkage model in an agile manner. Here we provide an overview of the application of data linkage models for strategies in combating COVID-19 in Taiwan, including NHIS centralized data linkage systems and "from border to community" information-driven data linkage systems during the COVID-19 pandemic. Furthermore, we discuss the dual role of digital technologies in being an "enabler" and a "driver" in early disease prevention. Lastly, Taiwan's experience in applying digital technology to enhance the control of COVID-19 potentially highlights lessons learned and opportunities for other countries to handle the COVID-19 situation better.

(J Med Internet Res 2021;23(5):e24294) doi:10.2196/24294

KEYWORDS

COVID-19; data linkage; digital health; digital technology; infectious disease; management; National Health Insurance System; prevention; spread; Taiwan

Introduction

Worldwide, there have been 60 million confirmed cases of COVID-19 and 1.4 million confirmed deaths as of November 30, 2020, and the numbers are still rising to date. When China reported its first COVID-19 outbreak, an early study conducted by Johns Hopkins University in January had forecasted the number of imported cases arriving at each airport inside and outside mainland China, and reported that Taiwan would be the second most affected country by COVID-19 owing to the close economic and transportation links with mainland China [1]. However, Taiwan has defied those expectations. Until November 30, 2020, Taiwan reported only 679 COVID-19 cases, 7 confirmed deaths, and only a 1.03% case fatality rate. Taiwan has been performing well in combating COVID-19 thus far; this may be attributed to the hard lessons learned from the severe acute respiratory syndrome (SARS) outbreak in 2003. The SARS experience had prepared the government and citizens of Taiwan

XSL•FO
RenderX

to respond to the COVID-19 pandemic more promptly and cautiously. Countries with a prior experience with SARS prevention (with over 100 confirmed cases and a >10% case fatality rate) also performed well in combating the current COVID-19 outbreak [2,3].

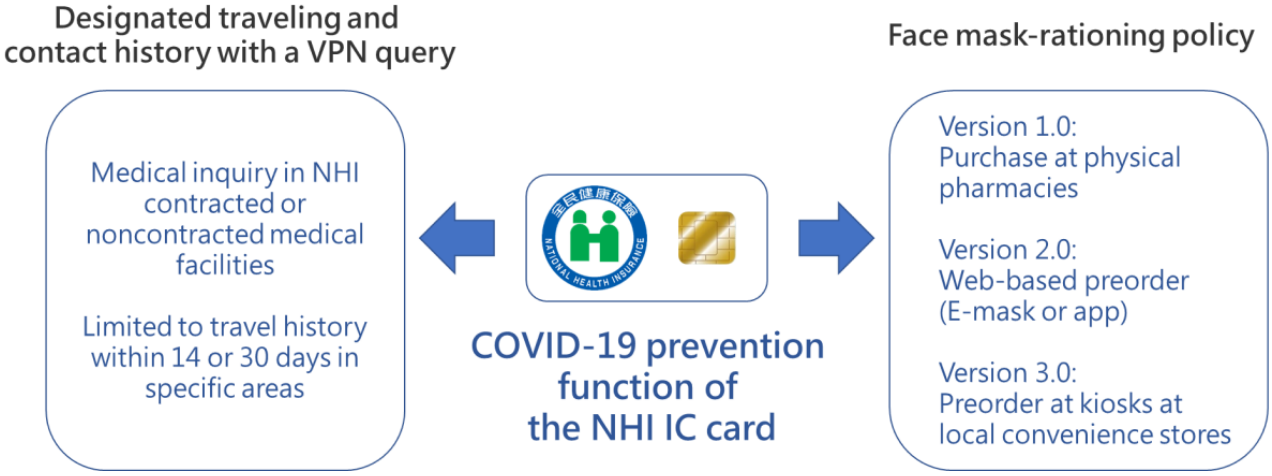
During the SARS outbreak of 2003 in Taiwan, the infected persons did not disclose their history of travel to outbreak areas in order to escape the quarantine; this was a loophole in border control. Furthermore, there was no regulation to prevent potentially infected people from entering hospitals without declaring their risk, resulting in the shutdown of a regional hospital and partial service closure at a medical center [4,5]. Furthermore, the face mask supplies in the market decreased and caused panic related to the purchase of face masks in the community. The devastating experience with the SARS outbreak had led Taiwan to be extremely cautious when China initially announced its COVID-19 outbreak on January 20, 2020. Taiwan quickly established the Central Epidemic Command Center (CECC). The well-developed infrastructure of the CECC has allowed for quick strategic planning, epidemic analysis, and disease prevention, as well as unifying medical resources to successfully contain the outbreak and minimize infections [6]. Meanwhile, digital technology (DT) has played an essential role in combating the COVID-19 pandemic (especially considering its dual roles as an "enabler" and a "driver") and provided tremendous help in containing SARS-CoV-2, which was not well-exploited during the SARS pandemic.

During the current COVID-19 pandemic, all countries worldwide deployed public health strategies including border control, control of community transmission, enhancement of personal hygiene, and prevention of nosocomial infection, and Taiwan was no exception. Here we elucidate how data linkage was used to strengthen disease prevention and to explore the role of DT in public health and disease control strategies. Based on Taiwan's successful experience with COVID-19 control, 2 major data linkage models are proposed herein for COVID-19 control. Furthermore, we have discussed the application of data linkage models.

Applications of the Current National Health Insurance System to Enhance COVID-19 Prevention and Control

Taiwan's National Health Insurance System (NHIS) was initially rolled out in 1995, with the National Health Insurance Administration (NHIA) being the only insurer. Since the NHIS was completely information-driven, 90% of the hospitals and clinics had used the electronic reporting systems to claim medical expenses. Currently, these services are fully automated and have been running smoothly for years. For security reasons, a virtual private network is used to strengthen data security while the information is being processed on the internet. Once the NHIA had collected, reported, and complied all medical expenses into the "centralized database," big data were used to automatically review and process all medical claims received from hospitals and clinics. As a benefit, Taiwan's NHIS has the lowest administrative cost worldwide [7,8]. In addition, NHIA issues National Health Insurance IC cards (NHI IC cards) as an insurance certificate. In general, Taiwanese citizens present the NHI IC card during medical visits. The NHI IC card enables physicians to obtain the most recent medical records of the patient. Moreover, detailed past medical records and medication use can also be accessed by linking this information with a medical cloud information exchange system, the MediCloud. Under the NHIS, two important pandemic prevention features were established: travel history tracking and the mask-rationing plan for purchase (Figure 1).

Figure 1. Diagrammatic representation of the NHI IC card usage in COVID-19 prevention policies. NHI: National Health Insurance, VPN: virtual private network.



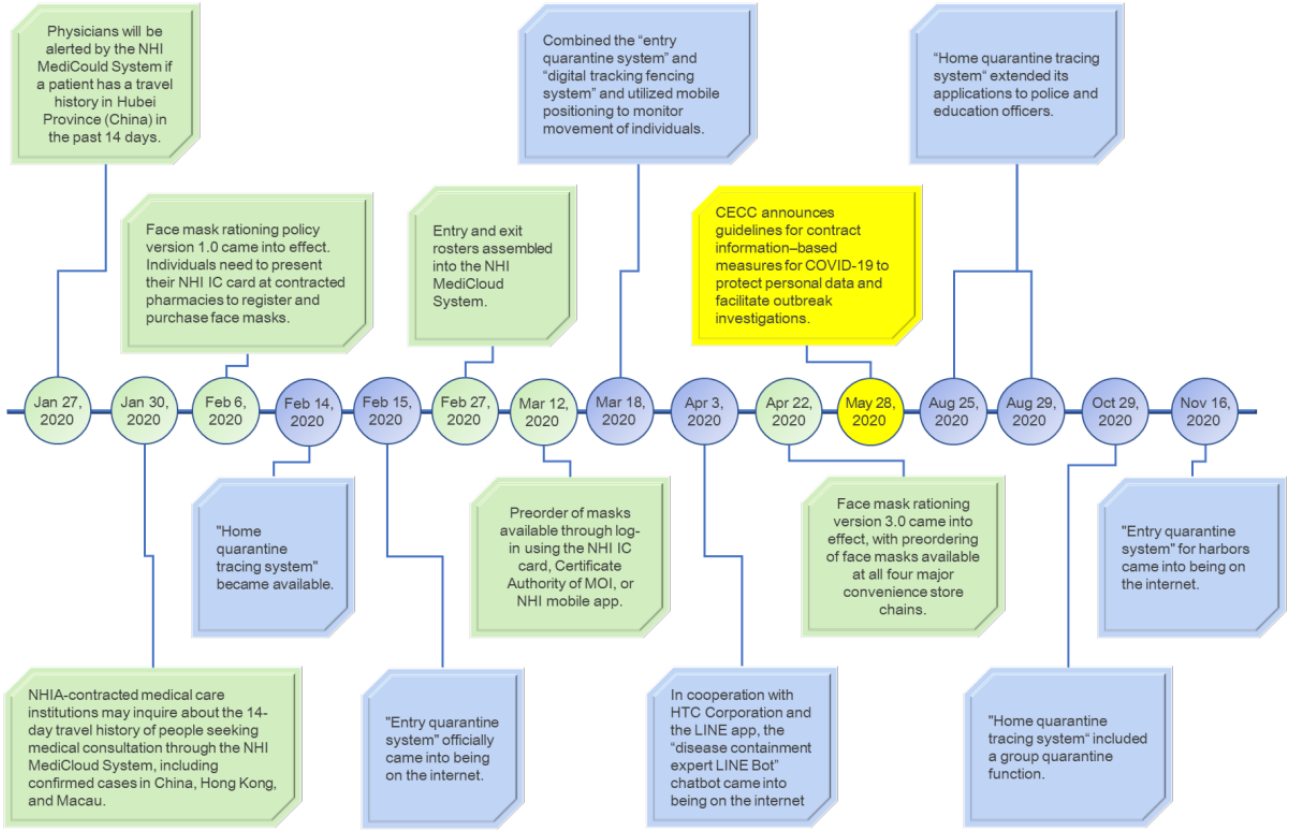
Travel History Tracking

At the early stage of CECC establishment, the presence of fever, cough, as well as other respiratory symptoms, travel history, occupation, contact history, and cluster exposure were used as the criteria for case reporting and screening. In addition to the existing pandemic reporting and screening procedures, toward the end of January 2020 (also known as Lunar New Year for Asian-Pacific regions), the CECC decided to utilize the MediCloud system from NHIS to provide inquiry services for travel history. MediCloud system (previously known as "PharmaCloud") had been expanding its services and functions

XSL•FO
RenderX

to provide patient information (ie, disease diagnosis, examinations, laboratory testing, and medication) to hospitals and clinics since its initiation in 2015. After years of promotion and system advancement, the usage of MediCloud is currently over 99%, and MediCloud has become the essential platform for information sharing among medical institutions. Once interface linkage was established between the National Immigration Agency (part of the Ministry of the Interior of Taiwan) and the NHIA, the travel information database was integrated with medical records within just a few days. Now the NHIA has been able to provide inquiry services for travel history among infected countries or regions. In response to the progression of the COVID-19 pandemic, the features of querying wider ranges of travel history and occupations were included (Figure 2). These features contribute greatly to the prevention of nosocomial infections and enable timely screening to prevent COVID-19 outbreaks. Front-line medical workers are able to identify potentially infected individuals on the basis of their travel history and have reported symptoms with the aid of real-time alerts from the integrated information system and taken appropriate steps to adequately protect themselves from COVID-19.

Figure 2. Timeline of technology usage in COVID-19 prevention in Taiwan. Green blocks represent the launching date of the travel history inquiry system and the name-based mask-rationing system, blue blocks represent the procedures and services for border control and the home quarantine system, and the yellow block represents the date when CECC announced guidelines for contact information–based measures and data collection. CECC: Central Epidemic Command Center, MOI: Ministry of the Interior, NHI: National health Insurance, NHIA: National Health Insurance Administration.



Name-Based Mask-Rationing System for Purchase

At the beginning of February 2020, the CECC announced that the daily production of face masks was approximately 2 million, which was markedly lesser than the total population of Taiwan (approximately 23 million). Meanwhile, China also proceeded with purchasing large amounts of face masks in response to the COVID-19 pandemic. Since the people of Taiwan have had experiences with the SARS pandemic, people in some areas started stocking up face masks at home to avoid shortage [9]. In response to the face mask shortfall, the government began to increase face mask production to ensure all citizens have equal access to high-quality face masks. The policy that the government deployed, the "name-based mask-rationing system," inevitably ensured fair distribution of face masks in the community. Since all insurees (including legal foreign workers in Taiwan) obtain an NHI IC card, the CECC instructed the NHIS to contract pharmacies to link with the MediCloud system such that the policy of "three face masks per week per individual" can be executed for all insured individuals. The policy utilized the information that the NHI IC card provides to verify and control the allocation of face masks to ensure that everyone has access to face masks for basic protection against COVID-19. Furthermore, at the beginning of March, the CECC had announced a payment mechanism where the public can utilize kiosks to reserve face masks, pay the fees, and collect them at either local convenience stores or pharmacies per their preference. Local manufacturers in Taiwan have been able to increase face mask production; hence, the purchasing policy was modified by the CECC to "nine face masks every two weeks per individual" in early May. Starting on June 1, 2020, the public was allowed to make more purchases if needed, other than the primary rationing.

XSL•FO
RenderX

The current infrastructure of the NHIS allows the aforementioned functions such as travel history tracking and executing a "name-based mask-rationing system" policy; thus, all pandemic prevention strategies postulated by the CECC can be implemented in a timely manner. The establishment of the NHI IC card, MediCloud, and a virtual private network were mostly completed after the SARS outbreak in 2003. The original purpose was not pandemic control, but rather for Taiwan's citizens to utilize medical services at their convenience. The original purpose of establishing the NHI IC card was to avoid the misuse of medical services by the public and to enhance medication safety by using information technologies to share information among hospitals, clinics, and pharmacies rather than for pandemic control. Meanwhile, this "centralized" data linkage structure has become the key enabler for the government to implement pandemic prevention strategies.

Establishment of a Pandemic Prevention System from the Border to the Community

Toward the end of January 2020, the CECC decided to instruct passengers arriving from China, Hong Kong, and Macau to follow home quarantine procedures. The procedure required passengers to receive a home quarantine notice once they entered Taiwan (both by air and sea). Both the civil affairs department and the health department could be informed with lists of people who require home quarantine in their responsible areas. The civil affairs personnel would then monitor those who are in home quarantine with regard to basic living and health needs, and the health department would arrange for medical treatments if needed.

Information-Driven System Planning

Owing to the lack of assistance from DT, the manpower required to carry out those surveillance duties was tremendous and ineffective in the beginning. For instance, the airport staff needed to identify those who require home quarantine from among 70,000 passengers entering Taiwan each day. Therefore, airports needed to recruit a large number of health officers or temporary assistants to carry out verification duties. According to the infectious disease prevention and treatment law, the duty of performing home quarantine is fulfilled by civil affairs personnel of the Ministry of Home Affairs and the local government. Health officers who are responsible for COVID-19 testing at the airport collected all the information on paper, performed web-based data entry, and the information was sent out from the health department to the Ministry of Home Affairs and then passed on to the local government. The local government passed on the information regarding the results of COVID-19 testing to the chief of the village, who in turn would check and assist those who are in home quarantine. Those who violate home quarantine procedures are subject to administrative penalties by the health administrative system. Many government employees are under extreme pressure with regard to mishandling of COVID-19–related tasks and the creation of loopholes in the disease prevention system during the pandemic. The delay in information received by the local government employees, misinformation regarding personal data (such as telephone numbers and addresses), quality of information, and efficiency may all contribute to the pressure felt by the government employees.

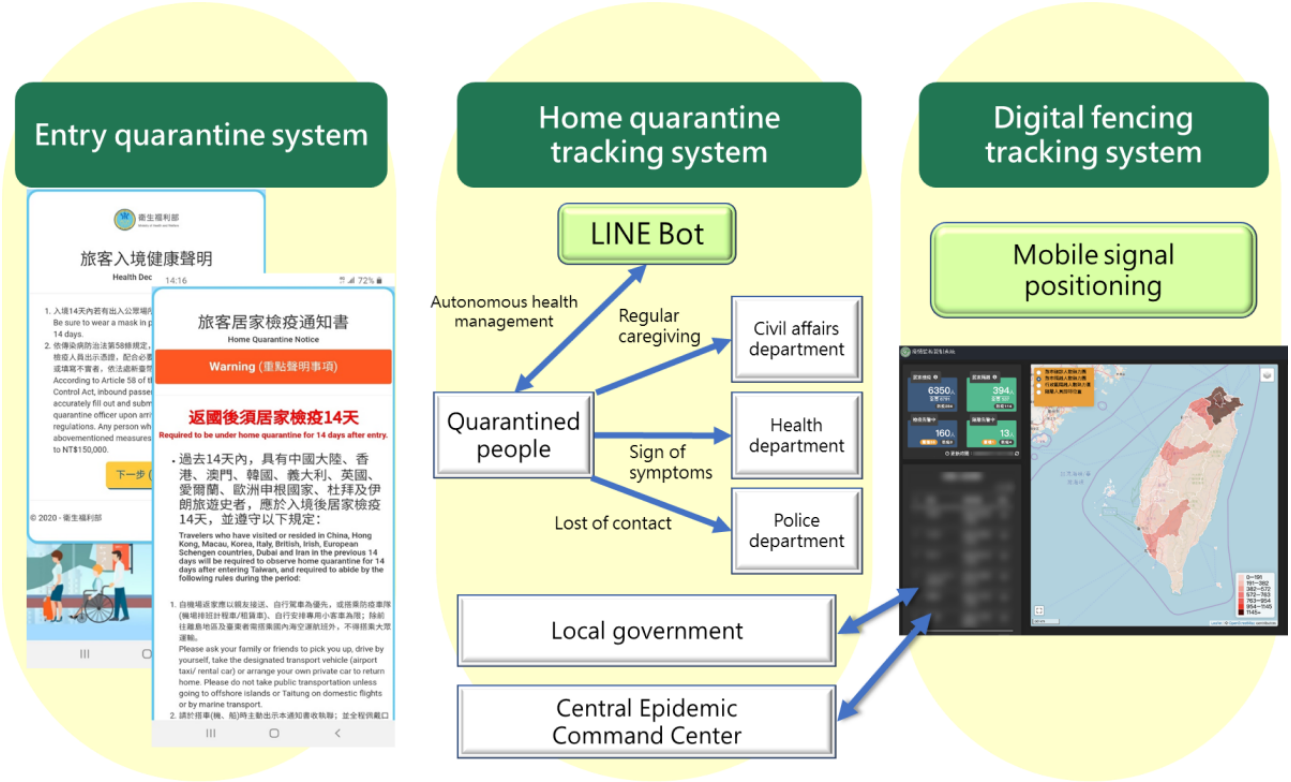
By the end of January 2020, the CECC had begun to improve data quality and efficiency and attempted to improve the sharing of correct personal information such as residential addresses, telephone numbers, cellphone numbers, and the start date of home quarantine among different government departments. As an information center, the CECC has established both the information flow and a relevant application system to allow for the circulation of information from the border to the community. Because 80% of the people of Taiwan are smartphone users, smartphones were being used as the main tool to keep track of individuals. A mobile app was not considered a viable option regardless of the smartphone's operating system (iOS or Android) since it may take considerable time for approval. Nevertheless, more time may be required for program correction. Therefore, the option of using a mobile app was eliminated as it was time-consuming during the critical outbreak.

Establishment of the Information-Driven System

The information-driven system was ready for launch after 2 weeks of brainstorming ideas to the completion of the entire system. In mid-February 2020, the information system became available and allowed data linkage among border control and local government services with the Advance Passenger Information System from the National Immigration Agency and Household Registration Information from the department of household affairs (also part of the Ministry of the Interior of Taiwan) for the management of home quarantine cases (Figure 3). For instance, the information system would ask all inbound passengers to use smartphones to scan a specific QR code prior to their departure and fill out personal information, health status, travel history, cellphone number, and public transportation needs. Before the aircraft reaches the gate, passengers will receive text messages regarding information on home quarantine notice and health declaration certificates. If passengers report symptoms related to COVID-19 under health status, the health personnel would proceed to test them immediately and arrange for medical treatment. If passengers indicate the need for public transportation, arrangements would be made. At present, there are on average 40,000 travelers arriving in Taiwan every day, and our information system has been functioning well in entry control.

XSL•FO
RenderX

Figure 3. Diagrammatic representation of the border control and home quarantine systems.



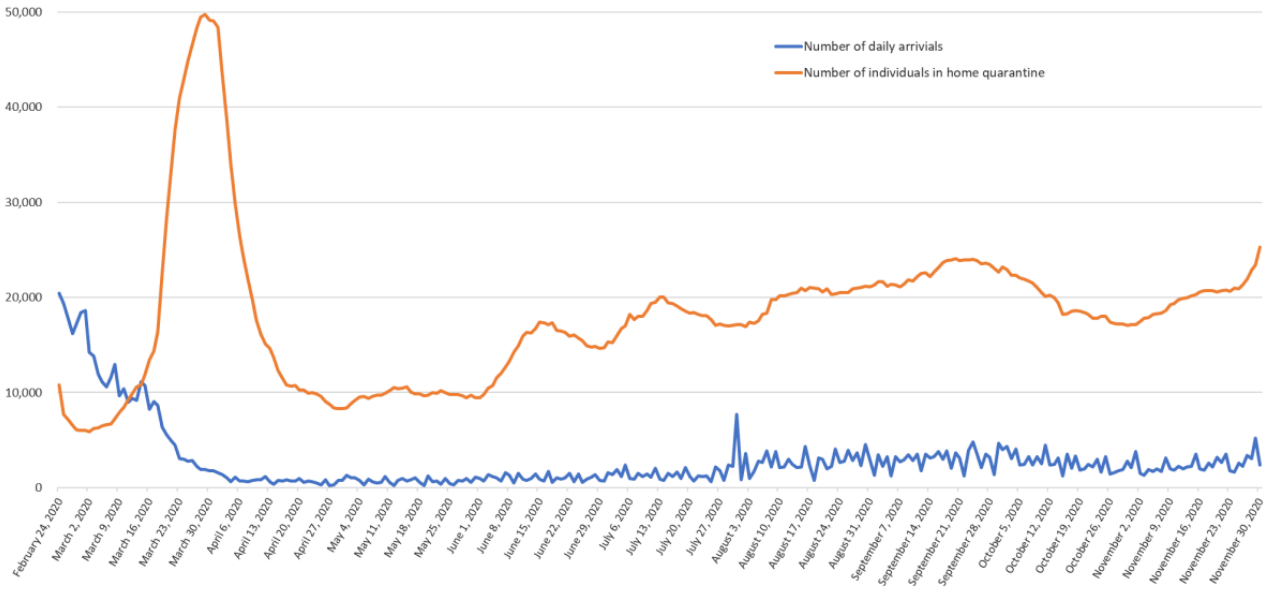
Immigration information can be delivered in a timely manner to the home quarantine management system, which is based on the regional joint defense concept to directly assign people who require home quarantine to the village, which prevents delays in information transfer. Initially, when the system was established, issues such as the submission of incorrect contact information were reported. The village-based civil affairs personnel (such as village officials) are responsible for confirming whether the individual has returned to the correct home quarantine location; if not, all the other possible addresses or telephone numbers may be used to track the individual down. If needed, the civil affairs personnel would check on the individual in person to ensure the individual stays at the designated quarantine location. The civil affairs personnel from different villages would also assist one another to carry out the surveillance duty. During the 14-day home quarantine period, the civil affairs personnel would provide home care services to the quarantined individuals on the basis of the address and contact number that appears in the system. Home care services provided by the civil affairs personnel are often delivered to the individual through smartphones, which will allow tracing of records for future reference. If there is an incidence of violation of home quarantine, the CECC would coordinate with telephone service providers to send out timely text messages to remind the concerned individual to return to their designated home quarantine location.

Merging of Information-Driven Systems and Instant Messaging Apps

As the system functions better each day and the accuracy of information continues to improve, the initiation of complete border control still remains a challenge. On March 19, 2020, the CECC announced that all travelers entering Taiwan will need to fulfill the home quarantine requirement. Although on average only approximately 7000 travelers entered Taiwan each day, the number of home quarantine cases still increased significantly (maximum 50,000 people/day to be monitored), and this certainly adds tremendous pressure on the local civil affairs personnel (Figure 4). The CECC used the media as a channel to advocate and invite the citizens to use the instant messaging app LINE and add the LINE Bot as a friend. On obtaining a consensus from home quarantine cases, the local civil affairs personnel could use the LINE app to follow up on home-quarantined individuals with regard to self-health management and personal needs on a daily basis. According to a survey performed in mid-May 2020, 7100 home-quarantined people in Taiwan who used the LINE app were sampled. The data revealed that chatting with LINE Bot and chatting with local civil affairs personnel scored 8.68 and 8.56, respectively, on a scale of 10 (0=completely unsatisfactory and 10=very satisfied).

XSL•FO
RenderX

Figure 4. Number of daily arrivals and home quarantine management in Taiwan (data were collected until November 30, 2020).



Until June 30, 2020, form among 447 confirmed cases of COVID-19, 356 (80%) were imported, and only 55 (12%) local cases and 36 (8%) confirmed cases were from navy clusters. For the imported cases, 141 infected individuals tested positive on arrival at the Taiwan border, 150 infected individuals tested positive during the 14-day quarantine, and the remaining 65 individuals tested positive during cluster quarantine, home quarantine, or isolation (Table 1). The information-based disease control system was the main "driver" to successfully block the spread of COVID-19 at the border and away from the community. The local rate of penalization for violating the home quarantine policy was lower than 0.5% (n=720/175,720 people).

Table 1. Source of confirmed COVID-19 cases in Taiwan as on November 30, 2020.

Records	Quantity
Tests performed, n	110,685
Confirmed cases (positive rate), n (%)[a]	679 (0.6)
Imported cases	588 (86.6)
Immigration testing	232 (39.5)
During home quarantine	207 (35.2)
Others	149 (25.3)
Locally acquired infections	55 (8.1)
Navy crew members aboard the Panshi Fast Combat Support Ship	36 (5.3)

[a]Imported cases, locally acquired infections, and navy crew members aboard the Panshi Fast Combat Support Ship are calculated relative to the number of confirmed cases.

Recommendation and Discussion

From the perspective of using data linkage to enhance pandemic control, the success of COVID-19 control in Taiwan was analyzed and could be attributed to 2 main data linkage models. These 2 main models enhance and improve 4 key public health strategies in Taiwan, including border control, prevention of community outbreaks, enhancement of personal hygiene, and prevention of nosocomial infections. Through data linkage between the NHIS and other departments, Taiwan was able to secure the supply of face masks and avoid nosocomial infections, which have been reported in previous studies [10,11]. The efficiency of the "centralized" data linkage model can be optimized as long as the integrity of the database is sufficient. Moreover, the range of travel history and the traveler's identity can be expanded as the pandemic progresses. The amount of allowance to purchase face masks and participating vendors can also be adjusted as needed. Overall, it is quite feasible to have the aforementioned flexibilities in a country such as Taiwan where the NHIS has demonstrated great functionality and performance. Unlike Taiwan, the United States has multiple insurance systems. Nevertheless, it is still feasible to use existing web-based medical networks to create similar data linkages through additional effort. From the border to the community, the information-oriented prevention system is considered a type of "downstream supply chain" of data linkage, and the key to

XSL•FO
RenderX

its success is attributed to the accuracy and timeliness of data collection [6,12]. For instance, for people who require home quarantine, within the first hour or 2 hours of arrival, the information should be sent directly to the civil affairs personnel of the appropriate residential area. This process can significantly ensure that information is received in a timely manner and can avoid any flaws in disease prevention. It is essential to implement the correct system at the right time to reinforce border management and prevent community outbreaks. Therefore, using the agile approach instead of the traditional approach in building the correct system is preferable during the COVID-19 outbreak.

Similar to other countries that rely heavily on information technologies, it is essential to establish a balance between disease prevention and personal data protection. However, the public generally agrees with and follows the necessary pandemic control policies despite the concern of personal data protection. Unlike Taiwan, South Korea has implemented a law allowing the government to reduce the level of personal data protection during the critical phase of the pandemic [13]. The government of Taiwan applies the current Infectious Disease Control Act and Personal Data Protection Act and openly discloses the prevention measures that are required to be followed by all travelers on entering Taiwan. On May 28, 2020, the CECC announced guidelines for contact information–based measures and data collection in response to the COVID-19 outbreak to ensure personal data protection and facilitate outbreak investigations. On the other hand, Taiwan chose to use a less privacy-sensitive DT approach, such as using smartphone-based station tracking instead of real-time GPS technology to monitor people who are in home quarantine or home isolation. Furthermore, a proposal has been approved, endorsed, and authorized by the legislature to eliminate all the unnecessary electronic records and linkages after the pandemic, which can also be adopted by other countries.

Currently, Taiwan has been free from local COVID-19 transmission for more than 200 days. The CECC has cautiously reopened the economy in a stepwise manner. As the restrictions on social activities are being lifted gradually, the “from border to community” information-driven prevention system continues to strengthen border control, ensuring that imported cases can be isolated immediately to minimize contact with others and information on the number of screenings performed and the number of confirmed cases can be compiled at the airport on a daily basis. Furthermore, the system can also help analyze and understand the changes in the number of confirmed cases overseas once border restrictions are lifted and adjustments are made to border control policies accordingly. Furthermore, in the post–COVID-19 era, new technology-driven practices will now form a part of “the new normal.” The government and organizations need to adopt new DT systems during the pandemic and devise methods to normalize these new practices. These new technology-driven work practices are usually implemented during the most severe times under high-pressure conditions and often without former experience or training. Taiwan’s experience can therefore be considered a valuable reference for other countries to further the current understanding of how DT can be embedded within government practices and form the “new normal” in the post–COVID-19 era.

In conclusion, the integrity, accuracy, and timeliness of data linkage and DT infrastructure ensure that essential public health interventions, such as border control, quarantine, case detection, contact tracing, and universal surgical mask–wearing can be effectively implemented and become the foundation of the highly successful COVID-19 pandemic response in Taiwan.

Acknowledgments

We appreciate the support from the Department of Information Management and Centers for Disease Control at the Ministry of Health and Welfare, Department of Cyber Security of the Executive Yuan, and the National Immigration Agency at the Ministry of the Interior of Taiwan.

Authors' Contributions

All authors contributed substantially to this study. IMP and HYC conceived and designed the study, and contributed to the manuscript equally. FCW, HWJ, HCW, and YLH prepared the figures and tables. WCL, FCW, and LYL drafted the manuscript. All authors contributed to the interpretation of data, critically revised the manuscript, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Modeling the Spread of 2019-nCoV. Johns Hopkins Whiting School of Engineering: Center for Systems Science and Engineering. 2020 Jan 26. URL: https://systems.jhu.edu/wp-content/uploads/2020/01/Gardner-JHU_nCoV-Modeling-Report_Jan-26.pdf [accessed 2020-05-21]
2. Summary table of SARS cases by country, 1 November 2002–7 August 2003. World Health Organization: Weekly Epidemiological Record. 2003 Aug 29. URL: https://apps.who.int/iris/bitstream/handle/10665/232250/WER7835_310-311.PDF [accessed 2021-05-03]

XSL•FO
RenderX

3. Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). World Health Organization. 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/70863/WHO_CDS_CSR_GAR_2003.11_eng.pdf [accessed 2021-05-03]
4. Hsieh Y, King C, Chen CWS, Ho M, Lee J, Liu F, et al. Quarantine for SARS, Taiwan. Emerg Infect Dis 2005 Feb;11(2):278-282 [FREE Full text] [doi: 10.3201/eid1102.040190] [Medline: 15752447]
5. Fan Y, Chen MC. The weakness of a post-authoritarian democratic society: reflections upon Taiwan's societal crisis during the SARS outbreak. In: Davis D, Siu HF, editors. SARS: Reception and Interpretation in Three Chinese Cities. London: Routledge; 2006:146-164.
6. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
7. Wu T, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. London J Prim Care (Abingdon) 2010 Dec;3(2):115-119 [FREE Full text] [doi: 10.1080/17571472.2010.11493315] [Medline: 25949636]
8. Mathauer I, Nicolle E. A global overview of health insurance administrative costs: what are the reasons for variations found? Health Policy 2011 Oct;102(2-3):235-246. [doi: 10.1016/j.healthpol.2011.07.009] [Medline: 21880390]
9. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
10. Schwartz J, King CC, Yen MY. Protecting Healthcare Workers During the Coronavirus Disease 2019 (COVID-19) Outbreak: Lessons From Taiwan's Severe Acute Respiratory Syndrome Response. Clin Infect Dis 2020 Jul 28;71(15):858-860 [FREE Full text] [doi: 10.1093/cid/ciaa255] [Medline: 32166318]
11. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
12. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. JAMA Intern Med 2020 Sep 01;180(9):1156-1163 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]
13. Chen C, Jyan H, Chien S, Jen H, Hsu C, Lee P, et al. Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]

Abbreviations

CECC: Central Epidemic Command Center
DT: digital technology
NHIA: National Health Insurance Administration
NHI IC card: National Health Insurance IC card
NHIS: National Health Insurance System
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 13.09.20; peer-reviewed by CT Fang, X Li, TH Kwan; comments to author 18.11.20; revised version received 08.12.20; accepted 15.04.21; published 07.05.21.

Please cite as:
Lo WC, Wang FC, Lin LY, Jyan HW, Wu HC, Huang YL, Parng IM, Chiou HY
Enhancing Data Linkage to Break the Chain of COVID-19 Spread: The Taiwan Experience
J Med Internet Res 2021;23(5):e24294
URL: https://www.jmir.org/2021/5/e24294
doi:10.2196/24294
PMID:33882019

©Wei-Cheng Lo, Fu-Chung Wang, Li-Yin Lin, Hong-Wei Jyan, Hsuan-Chien Wu, Yao-Liang Huang, I-Ming Parng, Hung-Yi Chiou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.05.2021. 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

Understanding Public Perceptions of COVID-19 Contact Tracing Apps: Artificial Intelligence–Enabled Social Media Analysis

Kathrin Cresswell[1], BSc, MSc, PhD; Ahsen Tahir[2,3], BSc, MSc, PhD; Zakariya Sheikh[4]; Zain Hussain[4], BSc, MBChB; Andrés Domínguez Hernández[5], BSc, MSc, PhD; Ewen Harrison[1], MB, ChB, MSc, PhD, FRCS; Robin Williams[6], MA, MSc, PhD; Aziz Sheikh[1], BSc, MBBS, MSc, MD; Amir Hussain[2], BEng, PhD

[1]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[2]School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

[3]University of Engineering and Technology, Lahore, Pakistan

[4]Edinburgh Medical School, The University of Edinburgh, Edinburgh, United Kingdom

[5]Department of Computer Science, University of Bristol, Bristol, United Kingdom

[6]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Kathrin Cresswell, BSc, MSc, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 (0)131 651 4151
Email: kathrin.cresswell@ed.ac.uk

Abstract

Background: The emergence of SARS-CoV-2 in late 2019 and its subsequent spread worldwide continues to be a global health crisis. Many governments consider contact tracing of citizens through apps installed on mobile phones as a key mechanism to contain the spread of SARS-CoV-2.

Objective: In this study, we sought to explore the suitability of artificial intelligence (AI)–enabled social media analyses using Facebook and Twitter to understand public perceptions of COVID-19 contact tracing apps in the United Kingdom.

Methods: We extracted and analyzed over 10,000 relevant social media posts across an 8-month period, from March 1 to October 31, 2020. We used an initial filter with COVID-19–related keywords, which were predefined as part of an open Twitter-based COVID-19 dataset. We then applied a second filter using contract tracing app–related keywords and a geographical filter. We developed and utilized a hybrid, rule-based ensemble model, combining state-of-the-art lexicon rule-based and deep learning–based approaches.

Results: Overall, we observed 76% positive and 12% negative sentiments, with the majority of negative sentiments reported in the North of England. These sentiments varied over time, likely influenced by ongoing public debates around implementing app-based contact tracing by using a centralized model where data would be shared with the health service, compared with decentralized contact-tracing technology.

Conclusions: Variations in sentiments corroborate with ongoing debates surrounding the information governance of health-related information. AI-enabled social media analysis of public attitudes in health care can help facilitate the implementation of effective public health campaigns.

(J Med Internet Res 2021;23(5):e26618) doi:10.2196/26618

KEYWORDS

artificial intelligence; sentiment analysis; COVID-19; contact tracing; social media; perception; app; exploratory; suitability; AI; Facebook; Twitter; United Kingdom; sentiment; attitude; infodemiology; infoveillance

XSL•FO
RenderX

Introduction

SARS-CoV-2 is a major global threat to public health. Many governments worldwide are using mobile apps for contact tracing of citizens to contain the spread of SARS-CoV-2 [1]. The contact tracing applications used vary within and between countries, but they generally track movements of individuals and aggregate data to communicate to users when they may have been exposed to the virus, and if they need to be tested and/or self-isolated.

The higher the user base of contact tracing apps, the larger the anticipated impact on reducing the estimated effective reproduction number of the virus [2,3]. However, there is inconsistent and incomplete uptake of these technologies, especially in individualistic societies where intended users may perceive such apps to be of limited personal benefit or infringing on personal privacy and where the use of such apps is difficult to mandate [4,5]. To understand what public health measures may promote the use of contact tracing apps in the United Kingdom, policy makers need to understand the reasons why members of the public may be hesitant to use them. The empirical literature shows how attitudes of prospective users influenced by their individual backgrounds, technological features of the app in question, and perceived benefits and trade-offs can be a key barrier to the effective implementation of such apps (see Textbox 1) [6-9].

More than half of the world's population, and around two-thirds of the UK population, currently use social media platforms, with significantly increased engagement levels during the ongoing pandemic [10]. We here sought to explore how analyzing public attitudes on these forums can provide insights into low levels of adoption of contact tracing apps in the United Kingdom. Our recent work has demonstrated the value of artificial intelligence (AI)–based sentiment analysis of social media data [11].

Textbox 1. Factors likely to impact the use of contact tracing apps identified in the literature.

- **Individual backgrounds:** demographics, health status, involvement with COVID-19, and previous experiences (technology, health, and other factors associated with COVID-19).
- **Technological features:** effectiveness, privacy, security, cost, trust in system vendor, performance and reliability (eg, false-positive or false-negative alerts), compatibility with installed base, interoperability, and architecture of systems (eg, centralized vs decentralized).
- **Perceived benefits:** self, society, research, epidemiology, mediators (eg, if the user or others will take action based on risk, if they care about the environment or society, if they believe data will be used effectively), and tension between perceived benefit to the user and altruistic benefit to the whole population.
- **Perceived trade-offs, risks, and limiting factors:** data collected (who will know what about the user), risk of marginalization of certain demographic groups, level of control of who sees what aspects of data and who can retain it for how long, and transparency of app.

Methods

Data Sources

To assess the potential of an AI-based sentiment analysis to understand public views and concerns, we analyzed data from two popular social media platforms—Facebook and Twitter. Facebook posts were extracted using the CrowdTangle platform [12], and Twitter posts were extracted from the COVID-19 Twitter chatter dataset, constructed by Panacea Lab (using a publicly available Twitter application programming interface) [13]. English-language Facebook posts and tweets posted in the United Kingdom from March 1 to October 31, 2020, were extracted, and thematically filtered using a two-step process, for keywords related to both COVID-19 and contact tracing apps. The Twitter dataset was already filtered with predefined COVID-19–related keywords as detailed previously [13], and the same keywords were used to apply an initial filter to the Facebook data. The following contract tracing app–related keywords for the second step filtering were selected by our interdisciplinary team: "covid app," "tracing app," "contact tracing," "privacy," "security," "app security," "app privacy," "contain virus spread," "movement tracking," "tracking," and "surveillance." Following geographical filtering for the United Kingdom, a total of 2000 tweets and 8000 Facebook posts were obtained for the analysis.

Analysis

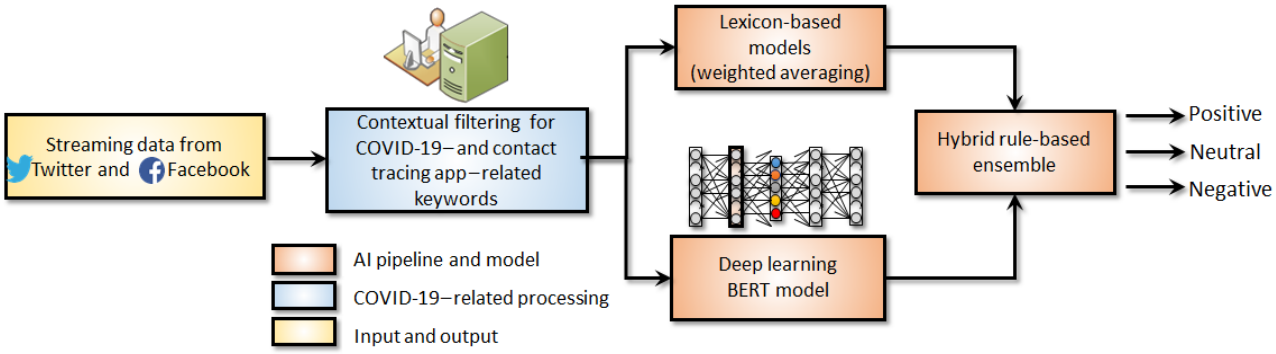
A novel ensemble-based AI model developed by the authors in a recent study [14] was adapted and utilized for this study, by combining lexicon rule-based and deep learning (DL)–based approaches. Specifically, an average-weighting ensemble [15] of lexicon-based models, including Valence Aware Dictionary and Sentiment Reasoner (VADER) [16] and TextBlob [17], was combined with a state-of-the-art DL-based model Bidirectional Encoder Representations from Transformers (BERT) [18]. The overall hybrid rule-based ensemble method is illustrated in Figure 1.

The hybrid rule-based ensemble model was hyperparameter tuned based on manual validation by using 1000 randomly selected posts and labelled by the team. Specificity and sensitivity analyses showed that the lexicon-based model demonstrated better performance accuracy for positive sentiment and that the DL-based BERT model demonstrated enhanced accuracy for neutral and negative sentiments, as shown in the confusion matrices (Multimedia Appendix 1). The TextBlob lexicon-based model utilizes a higher weight (0.52) than the VADER model, as it demonstrated marginally better accuracy (VADER × 0.48 + TextBlob × 0.52) for positive sentiments. For the ensemble, weighted averaging was utilized to combine the two lexicon-based models. The output of the weighted average lexicons was combined with the DL-based BERT model through a rule-based approach. For positive sentiments, the weighted average output of the lexicon models was selected as

XSL•FO
RenderX

the final output. For neutral and negative sentiments, the output of the DL-based BERT model was chosen as the final output. If-else logical constructs were used to determine the final sentiment class based on the base model outputs.

Figure 1. Hybrid rule-based ensemble model for sentiment classification of social media data. AI: artificial intelligence; BERT: Bidirectional Encoder Representations from Transformers.



Ethics

No ethical review was necessary since the data analyzed was fully in the public domain. A thorough assessment of the privacy risk to individuals posed by our research was conducted to ensure compliance with relevant sections of the General Data Protection Regulation (GDPR). We have also striven to comply with best practices for user protection to ensure no nonpublic material was included in our dataset.

Data Availability

The code and data used for data analysis and generation of figures are openly available on GitLab [19] for reproducibility and transparency of the analysis. Due to the computationally expensive nature of the code, we recommend using a high-performance computing resource.

Results

Figure 2 shows the average weekly sentiment trends on using contact tracing apps on Facebook and Twitter. The sentiments were computed by utilizing our ensemble-based AI model to predict the *overall* polarity of each post as positive, negative, or neutral. Due to the relatively small number, sentiments extracted from Facebook and Twitter posts were combined using weighted weekly averages. Sentiment word clouds were used to explore the topics being discussed at different time points of interest (Figure 2). Overall, the average positive sentiments were found to far outnumber the negative sentiments. We observed a six-fold difference between the sentiment type, with 76% positive and 12% negative sentiments. These were found to broadly corroborate with findings from independent surveys that show strong support for contact tracing apps. For example, an Ipsos MORI survey of 1983 UK adults conducted in May 2020 found that two-thirds (67%) of the population was in support and 12% was against the government's plans to use a smartphone app to track and trace new cases of COVID-19 [20]. However, the public's attitudes were found to be heavily influenced by demographics and the digital divide [12,21]. Several new survey studies are currently underway to explore public attitudes towards contact tracing apps.

Figure 2 illustrates some of the key events that are likely to have influenced changes in the public's sentiments toward using contact tracing apps. Following the successful implementation of contact tracing strategies in several East Asian countries and high-profile research reporting that app-based contact tracing was likely to be effective in curbing the spread of the virus [22-24], Apple and Google partnered on developing a decentralized COVID-19 contact tracing technology [25]. Shortly after, the UK government released contact tracing guidance with plans to develop a *home-grown* app and deploy it through a centralized model, implying that individual information would be shared with health services [26,27]. This model was controversial, primarily driven by public concerns around privacy and security of the app, its incompatibility with iPhones, and potential breaches of the Data Protection Law [28]. In June 2020, the centralized UK app was abandoned [29]. In August 2020, the UK government decided to implement the decentralized contact tracing system developed by Apple and Google [30], launching the app in September 2020 [31]. Since then, there have been concerns around high rates of false-positive COVID-19 cases reported via the app, hindering its uptake among the public [32].

We also carried out spatial mapping of the sentiments extracted from geo-tagged tweets to counties in the United Kingdom (see Figure 3). Most tweets had an overall positive sentiment toward COVID-19 contact tracing. The counties with the most positive sentiment included Lincolnshire, Norfolk, Nottinghamshire, Leicestershire, and Northamptonshire in England, and Stirling, Fife, Dumfries and Galloway, East Ayrshire, and West Lothian in Scotland. Counties with the most negative sentiment were Suffolk, Somerset, Devon, and North Yorkshire—all in England.

XSL•FO
RenderX

Figure 2. Average weekly sentiments extracted from Facebook and Twitter posts (combined) in the United Kingdom.

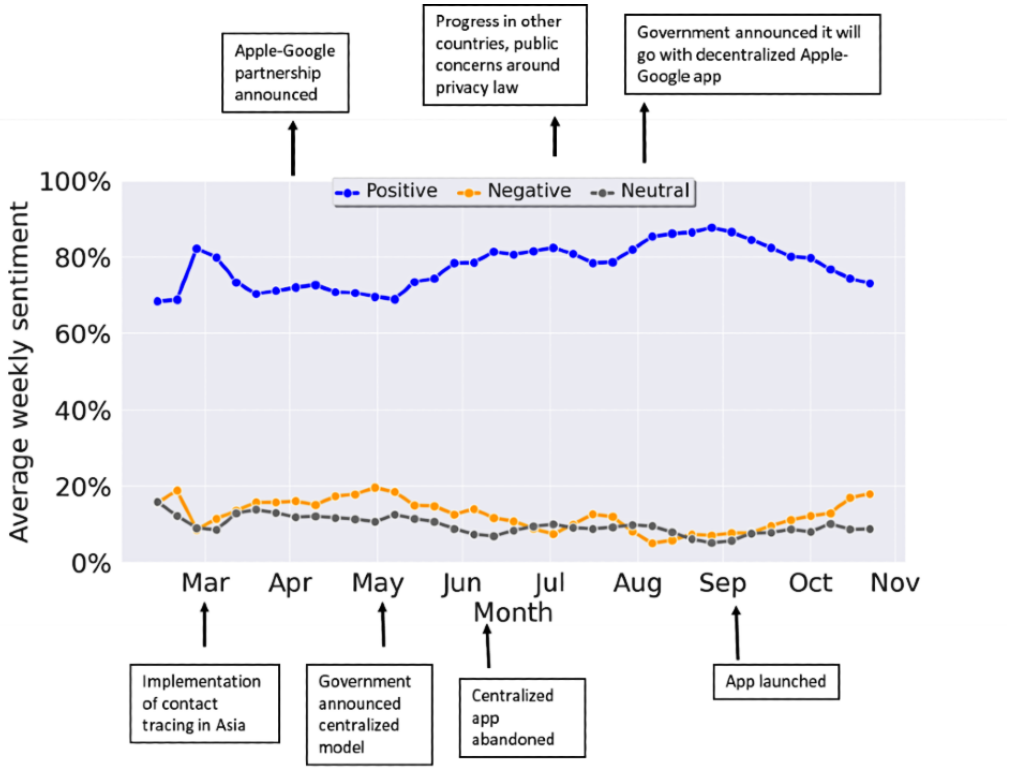
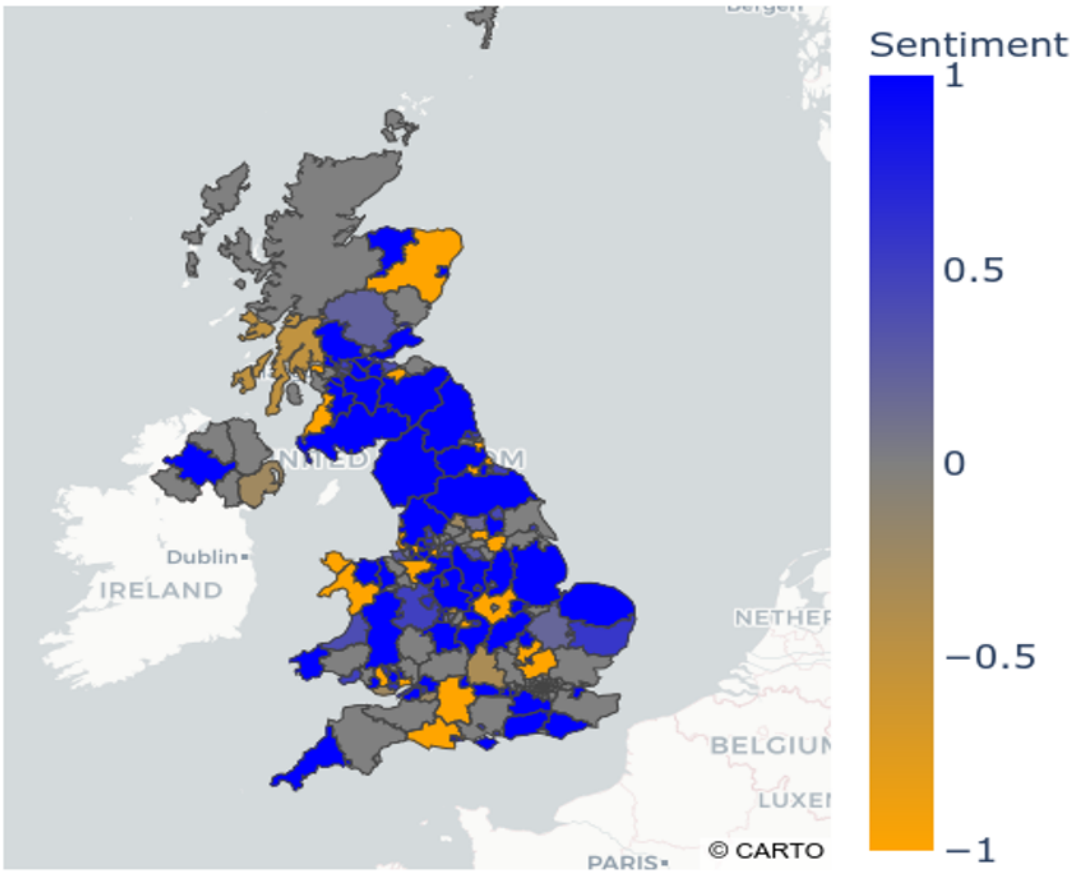


Figure 3. Spatial mapping of (averaged) UK public sentiment related to COVID-19 contact tracing apps on Twitter.



XSL•FO
RenderX

Discussion

Principal Findings

This study has provided insights into some of the underlying issues and concerns surrounding contact tracing apps, and particularly how decisions surrounding system architecture (eg, centralized vs decentralized) can influence public attitudes.

More fine-grained spatiotemporal variation in sentiments across the United Kingdom, and globally, should be further studied. Failure to pay timely attention to negative attitudes and resistance to emerging technologies in population health can have potentially damaging consequences, whereas an understanding of barriers can facilitate the development of effective interventions.

Limitations of our preliminary analysis include that social media platform users are largely not representative of the total UK population (eg, users are younger, more left-wing, and have higher incomes) [33]. Only about 15% of UK adults are regular Twitter users, with activity levels varying among passive and active users, and the demographics of users do not reflect the demographics of the general population [34]. It has also been shown that sentiments of tweets can vary across geographical locations and may therefore reflect the demographics of the population being studied [35]. Variations in sentiments can occur at an individual level, influenced by personal experiences and changes not necessarily related to the subject of the tweet [36]. The developed ensemble-based sentiment analysis approach can be optimized by using additional labelled data for transfer learning and fine-tuning the BERT base model. Adaptive neuro-fuzzy inferencing [37] can also be utilized to replace the current nonadaptive (weighted-average and logical rule-based) constructs.

The current focus on selected social media platforms, outstanding issues around the accuracy of AI techniques (eg, around deceptive language), and the limited number of tweets and Facebook posts due to the specific search strategy used, as well as given that only geo-tagged tweets were used, may limit the generalizability of these findings. Thus, there is a need for a more refined and comprehensive search strategy using multiple social media and web platforms, as well as linked analysis with data from external trustworthy sources such as polls, census, surveys, and clinical notes. Considering the limitations, the proposed approach should currently only be used alongside other methods to assess public sentiment, including public consultations, surveys, and qualitative studies.

In future, mobility trends [38] can be integrated into this approach and demographic determinants can also be inferred [39] and included to provide further insights. Furthermore, subgroup analysis could explore reasons for low uptake in certain populations or communities (eg, anti-vaxxers [40]), which can inform the design of vaccine deployment strategies, including public messaging campaigns. The approach may help promote a learning public health policy cycle, where ideas can be tested against public attitudes before they are implemented, thereby maximizing their effectiveness and real-world applicability [41].

Conclusions

We advocate for future work on AI-enabled social media analysis of public attitudes in health care, which we believe has the potential to help facilitate the implementation of effective public health campaigns. Through this preliminary analysis, we show how such innovative methods may complement findings using conventional methods to monitor public sentiments (eg, surveys) while also providing greater spatiotemporal granularity.

Acknowledgments

This work is supported by the Scottish Government Chief Scientist Office under its COVID-19 priority research program (grant COV/NAP/20/07), and by BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_19004]. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research, United Kingdom. AH is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) (grants EP/T021063/1 and EP/T024917/1). KC is supported by a Scottish Government Chief Scientist Office Research Grant exploring the potential of next-generation health information technology.

Conflicts of Interest

AS is a member of the Scottish Government's Chief Medical Officer's COVID-19 Advisory Group and the UK Government's New and Emerging Respiratory Virus Threats (NERVTAG) Risk Stratification Subgroup. The other authors have no conflicts to declare. The views expressed are those of the authors and does not represent the views of the Scottish or UK governments.

Multimedia Appendix 1
Normalized confusion matrices for VADER- and TextBlob-based lexicon models, deep learning-based BERT model, and the hybrid rule-based ensemble approach. BERT: Bidirectional Encoder-Representations from Transformers; VADER: Valence Aware Dictionary and Sentiment Reasoner.
[PNG File , 338 KB - jmir_v23i5e26618_fig.png]

References

XSL•FO
RenderX

1. Digital contact tracing can slow or even stop coronavirus transmission and ease us out of lockdown. University of Oxford. 2020 Apr 16. URL: https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown [accessed 2020-11-09]
2. Kretzschmar M, Rozhnova G, Bootsma M, van Boven M, van de Wijgert JHHM, Bonten M. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
3. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
4. Kubba S. The Importance of Culture in Societal Responses to COVID-19. Harvard Political Review. 2020 Oct 14. URL: https://harvardpolitics.com/culture-response-covid-19/ [accessed 2020-11-09]
5. Fendos J. How surveillance technology powered South Korea's COVID-19 response. Brookings. 2020 Apr 29. URL: https://www.brookings.edu/techstream/how-surveillance-technology-powered-south-koreas-covid-19-response/ [accessed 2020-11-09]
6. Redmiles EM. User Concerns & Tradeoffs in technology-facilitated contact tracing. Digit. Gov.: Res. Pract 2021 Jan 11;2(1):1-12 [FREE Full text] [doi: 10.1145/3428093]
7. Hargittai E, Redmiles E. Will Americans Be Willing to Install COVID-19 Tracking Apps? Scientific American. 2020 Apr 28. URL: https://tinyurl.com/3u87bkjr [accessed 2021-05-04]
8. Kaptchuk G, Hargittai E, Redmiles E. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. arXiv Preprint posted online on May 9, 2020.
9. Kent C. A comedy of errors: the UK's contact-tracing apps. Medical Device Network. 2020 Oct 6. URL: https://www.medicaldevice-network.com/features/uk-contact-tracing-app-problems/ [accessed 2021-03-16]
10. Perez S. Report: WhatsApp has seen a 40% increase in usage due to COVID-19 pandemic. Tech Crunch. 2020 Mar 26. URL: https://techcrunch.com/2020/03/26/report-whatsapp-has-seen-a-40-increase-in-usage-due-to-covid-19-pandemic/ [accessed 2020-11-09]
11. Al-Ghadir AI, Azmi AM, Hussain A. A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments. Information Fusion 2021 Mar;67:29-40. [doi: 10.1016/j.inffus.2020.10.003]
12. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
13. COVID-19 AI-powered dashboard for sentiment and opinion mining in social media platforms. COVID Tracker Cloud.: Edinburgh Napier University URL: http://covidtracker.cloud [accessed 2020-11-09]
14. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of UK and US public attitudes on Facebook and Twitter towards COVID-19 vaccinations. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
15. Yoo BH, Kim J, Lee B, Hoogenboom G, Kim KS. A surrogate weighted mean ensemble method to reduce the uncertainty at a regional scale for the calculation of potential evapotranspiration. Sci Rep 2020 Jan 21;10(1):870 [FREE Full text] [doi: 10.1038/s41598-020-57466-0] [Medline: 31964919]
16. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 May 16 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI URL: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf
17. Loria S, Keen P, Honnibal M, Yankovsky R, Karesh D, Dempsey E. TextBlob: simplified text processing. Secondary TextBlob: Simplified Text Processing. 2014. URL: https://textblob.readthedocs.io/en/dev/ [accessed 2020-08-10]
18. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. : Association for Computational Linguistics; 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 2019; Minneapolis, MN p. 4171-4186 URL: https://www.aclweb.org/anthology/N19-1423 [doi: 10.18653/v1/N19-1423]
19. Edinburgh Napier University - covid-app: Data and code for using artificial intelligence-enabled analysis of social media data to understand public perceptions of COVID-19 contact tracing apps. GitLab. URL: https://gitlab.com/edinburgh-napier-university/covid-app [accessed 2021-05-04]
20. Survey on COVID-19 track and trace smartphone app for The Health Foundation. IPSOS Mori. 2020 Jun 3. URL: https://www.ipsos.com/ipsos-mori/en-uk/survey-covid-19-track-and-trace-smartphone-app-health-foundation [accessed 2020-12-01]
21. Taylor E, Jackson J, Yesberg J, Pósch K. Coronavirus: survey reveals what the public wants from a contact-tracing app. The Conversation. 2020 May 15. URL: https://theconversation.com/coronavirus-survey-reveals-what-the-public-wants-from-a-contact-tracing-app-138574 [accessed 2020-12-01]

XSL•FO
RenderX

22. Kennedy J. What can the UK learn from South Korea's response to covid-19? Centre for Health and the Public Interest (CHPI). 2020 Apr 7. URL: https://chpi.org.uk/blog/what-can-the-uk-learn-from-south-koreas-response-to-covid-19/ [accessed 2020-11-09]
23. Utomo D. TraceTogether: Things you might not know about Singapore's contact tracing app. TimeOut. 2020 May 12. URL: https://www.timeout.com/singapore/news/tracetogether-things-you-might-not-know-about-singapores-contact-tracing-app-051220 [accessed 2020-11-09]
24. Kucharski A, Klepac P, Conlan A, Kissler S, Tang M, Fry H, CMMID COVID-19 working group. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis 2020 Oct;20(10):1151-1160 [FREE Full text] [doi: 10.1016/S1473-3099(20)30457-6] [Medline: 32559451]
25. Apple and Google partner on COVID-19 contact tracing technology. Apple Newsroom. 2020 Apr 10. URL: https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/ [accessed 2020-11-09]
26. NHS Test and Trace: how it works. Gov.UK. URL: https://www.gov.uk/guidance/nhs-test-and-trace-how-it-works [accessed 2020-11-09]
27. Coronavirus test, track and trace plan launched on Isle of Wight. Gov.UK. 2020 May 4. URL: https://www.gov.uk/government/news/coronavirus-test-track-and-trace-plan-launched-on-isle-of-wight [accessed 2020-11-09]
28. UK contact-tracing app could fall foul of privacy law, government told. URL: https://www.theguardian.com/world/2020/may/07/uk-coronavirus-contract-tracing-app-could-fall-foul-of-privacy-law-government-told [accessed 2020-11-09]
29. Murphy S, Sabbagh D, Hern A. Piloted in May, ditched in June: the failure of England's Covid-19 app. The Guardian. 2020 Jun 18. URL: https://www.theguardian.com/world/2020/jun/18/piloted-in-may-ditched-in-june-the-failure-of-englands-covid-19-app [accessed 2020-11-09]
30. Downey A. Covid-19: New trial for England’s revamped NHS contact-tracing app. Digital Health. 2020 Aug 13. URL: https://www.digitalhealth.net/2020/08/covid-19-new-trial-nhs-contact-tracing-app/ [accessed 2020-11-09]
31. Downey A. NHS Covid-19 contact-tracing app launched in England and Wales. Digital Health. URL: https://www.digitalhealth.net/2020/09/nhs-covid-19-contact-tracing-app-launched-england-wales/ [accessed 2020-11-09]
32. Manthorpe R. Coronavirus: NHS contact tracing app to stop sending 'false alarms' after update. Sky News. 2020 Oct 29. URL: https://news.sky.com/story/coronavirus-nhs-contact-tracing-app-to-stop-sending-false-alarms-after-update-12117882 [accessed 2020-11-09]
33. Blank G. The digital divide among Twitter users and its implications for social research. Social Science Computer Review 2016 Oct 12;35(6):679-697. [doi: 10.1177/0894439316671698]
34. Smith A, Brenner J. Twitter Use 2012. Pew Research Center. 2012 May 31. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2021-05-04]
35. 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]
36. 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]
37. Kour H, Manhas J, Sharma V. Usage and implementation of neuro-fuzzy systems for classification and prediction in the diagnosis of different types of medical disorders: a decade review. Artif Intell Rev 2020 Feb 04;53(7):4651-4706 [FREE Full text] [doi: 10.1007/s10462-020-09804-x]
38. Drake TM, Docherty AB, Weiser TG, Yule S, Sheikh A, Harrison EM. The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK. The Lancet Digital Health 2020 Aug;2(8):e385-e387. [doi: 10.1016/s2589-7500(20)30134-5]
39. Brandt J, Buckingham K, Buntain C, Anderson W, Ray S, Pool J, et al. Identifying social media user demographics and topic diversity with computational social science: a case study of a major international policy forum. J Comput Soc Sc 2020 Jan 07;3(1):167-188. [doi: 10.1007/s42001-019-00061-9]
40. Francia M, Gallinucci E, Golfarelli M. Social BI to understand the debate on vaccines on the Web and social media: unraveling the anti-, free, and pro-vax communities in Italy. Soc Netw Anal Min 2019 Aug 10;9:46. [doi: 10.1007/s13278-019-0590-x]
41. Hussain A, Sheikh A. Opportunities for artificial intelligence–enabled social media analysis of public attitudes toward Covid-19 vaccines. NEJM Catal Innov Care Deliv 2021 Feb 05;2(1). [doi: 10.1056/CAT.20.0649]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder-Representations from Transformers
DL: deep learning
GDPR: General Data Protection Regulation
VADER: Valence Aware Dictionary and Sentiment Reasoner

XSL•FO
RenderX

Edited by C Basch; submitted 19.12.20; peer-reviewed by J Li, R Gore, S Lalmuanawma; comments to author 15.03.21; revised version received 29.03.21; accepted 17.04.21; published 17.05.21.

Please cite as:
Cresswell K, Tahir A, Sheikh Z, Hussain Z, Domínguez Hernández A, Harrison E, Williams R, Sheikh A, Hussain A
Understanding Public Perceptions of COVID-19 Contact Tracing Apps: Artificial Intelligence–Enabled Social Media Analysis
J Med Internet Res 2021;23(5):e26618
URL: https://www.jmir.org/2021/5/e26618
doi:10.2196/26618
PMID:33939622

©Kathrin Cresswell, Ahsen Tahir, Zakariya Sheikh, Zain Hussain, Andrés Domínguez Hernández, Ewen Harrison, Robin Williams, Aziz Sheikh, Amir Hussain. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

An eHealth Framework for Managing Pediatric Growth Disorders and Growth Hormone Therapy

Paul Dimitri[1*], BSc, MBChB, PhD; Luis Fernandez-Luque[2*], PhD; Indraneel Banerjee[3], MD; Ignacio Bergadá[4], MD; Luis Eduardo Calliari[5], MD; Jovanna Dahlgren[6,7], MD; Antonio de Arriba[8], PhD, MD; Risto Lapatto[9], PhD, MD; Thomas Reinehr[10], MD; Senthil Senniappan[11], PhD, MD; Cécile Thomas-Teinturier[12], MD; Meng-Che Tsai[13], MD; Azriyanti Anuar Zaini[14], MBBS; Merat Bagha[15], MSEE; Ekaterina Koledova[16], PhD, MD

[1]The Academic Unit of Child Health, Sheffield Children's NHS Foundation Trust, Sheffield, United Kingdom

[2]Adhera Health Inc, Palo Alto, CA, United States

[3]Royal Manchester Children's Hospital, Manchester University Hospitals Foundation Trust, Manchester, United Kingdom

[4]Centro de Investigaciones Endocrinológicas "Dr. César Bergadá" (CEDIE), Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina

[5]Department of Paediatrics, Santa Casa de São Paulo School of Medical Sciences, São Paulo, Brazil

[6]Department of Pediatrics, Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden

[7]Department of Pediatrics, Institute of Clinical Sciences, University of Gothenburg, Gothenburg, Sweden

[8]Paediatric Endocrinology, Hospital Universitario Miguel Servet, Zaragoza, Spain

[9]New Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[10]Vestische Hospital for Children and Adolescents, University of Witten/Herdecke, Datteln, Germany

[11]Department of Paediatric Endocrinology, Alder Hey Children's Hospital, Liverpool, United Kingdom

[12]Department of Pediatric Endocrinology, Assistance Publique – Hôpitaux de Paris, Université Paris Saclay, Hôpital Bicetre, Le Kremlin Bicêtre, France

[13]Department of Pediatrics, National Cheng Kung University, Tainan, Taiwan

[14]Department of Pediatrics, University Malaya Medical Centre, Kuala Lumpur, Malaysia

[15]Tiba Medical Inc, Beaverton, OR, United States

[16]Global Medical Affairs, Cardiometabolic and Endocrinology, Merck KGaA, Darmstadt, Germany

[*]these authors contributed equally

Corresponding Author:
Paul Dimitri, BSc, MBChB, PhD
The Academic Unit of Child Health
Sheffield Children's NHS Foundation Trust
Damer St Building
Western Bank
Sheffield, S10 2TH
United Kingdom
Phone: 44 114 271 7118
Email: paul.dimitri@nhs.net

Abstract

Background: The use of technology to support health and health care has grown rapidly in the last decade across all ages and medical specialties. Newly developed eHealth tools are being implemented in long-term management of growth failure in children, a low prevalence pediatric endocrine disorder.

Objective: Our objective was to create a framework that can guide future implementation and research on the use of eHealth tools to support patients with growth disorders who require growth hormone therapy.

Methods: A total of 12 pediatric endocrinologists with experience in eHealth, from a wide geographical distribution, participated in a series of online discussions. We summarized the discussions of 3 workshops, conducted during 2020, on the use of eHealth in the management of growth disorders, which were structured to provide insights on existing challenges, opportunities, and solutions for the implementation of eHealth tools across the patient journey, from referral to the end of pediatric therapy.

Results: A total of 815 responses were collected from 2 questionnaire-based activities covering referral and diagnosis of growth disorders, and subsequent growth hormone therapy stages of the patient pathway, relating to physicians, nurses, and patients,

XSL•FO
RenderX

parents, or caregivers. We mapped the feedback from those discussions into a framework that we developed as a guide to integration of eHealth tools across the patient journey. Responses focused on improved clinical management, such as growth monitoring and automation of referral for early detection of growth disorders, which could trigger rapid evaluation and diagnosis. Patient support included the use of eHealth for enhanced patient and caregiver communication, better access to educational opportunities, and enhanced medical and psychological support during growth hormone therapy management. Given the potential availability of patient data from connected devices, artificial intelligence can be used to predict adherence and personalize patient support. Providing evidence to demonstrate the value and utility of eHealth tools will ensure that these tools are widely accepted, trusted, and used in clinical practice, but implementation issues (eg, adaptation to specific clinical settings) must be addressed.

Conclusions: The use of eHealth in growth hormone therapy has major potential to improve the management of growth disorders along the patient journey. Combining objective clinical information and patient adherence data is vital in supporting decision-making and the development of new eHealth tools. Involvement of clinicians and patients in the process of integrating such technologies into clinical practice is essential for implementation and developing evidence that eHealth tools can provide value across the patient pathway.

(J Med Internet Res 2021;23(5):e27446) doi:10.2196/27446

KEYWORDS

eHealth tools; pediatric growth disorders; referral and diagnosis; growth hormone therapy; adherence to treatment; workshop discussions; eHealth; pediatrics; growth failure; growth hormone

Introduction

eHealth Implementation

The implementation of eHealth solutions in clinical practice is growing rapidly [1]. Digitalization of the health sector is bringing many opportunities, such as the collection of patient-generated data [2] and the provision of health services over distance, also known as telehealth [3]. These technological advances are supporting a paradigm shift towards more integrated health care [4]. However, implementing technologies in the highly complex setting of health care delivery faces many challenges, which include both technical and human factors (eg, organizational issues, service integration, usability, legal frameworks, privacy) [5].

Opportunities of eHealth for Patient Empowerment in Growth Hormone Therapy

Pediatric endocrinology (the management of endocrine disorders in children) has adopted eHealth solutions for decades. For example, diabetes mellitus education uses game-based interventions [6,7], and more recent advanced technologies, such as augmented reality [8] and robotics [9], have been integrated into education and assistive care in diabetes mellitus. These approaches have traditionally looked at chronic conditions that require intensive lifestyle modifications, where lack of adherence to medication has a negative impact, with an increased risk of health complications and hospitalization.

Similarly, one of the areas that may benefit from implementation of novel eHealth tools is the long-term use of growth hormone therapy in the management of childhood growth failure. Growth hormone therapy is indicated for a heterogeneous set of growth disorders; in addition to growth hormone deficiency, approved pediatric indications include Turner syndrome, short stature after being born small for gestational age, Noonan syndrome, Prader-Willi syndrome, *SHOX* (short stature homeobox-containing gene) deficiency, chronic renal failure, and idiopathic short stature, although licensed indications vary by country and growth hormone formulations [10,11].

Importantly, growth hormone affects not only growth of children, but also metabolism, cardiovascular health, bone strength, and quality of life in the short and long term [12-14]. These effects continue into adult life, and a proportion of patients treated as children continue to require growth hormone therapy as adults. During the transition from pediatric to adult care, eHealth tools have been shown to be important in supporting patients to achieve positive outcomes and enhance ongoing engagement with health care [15,16].

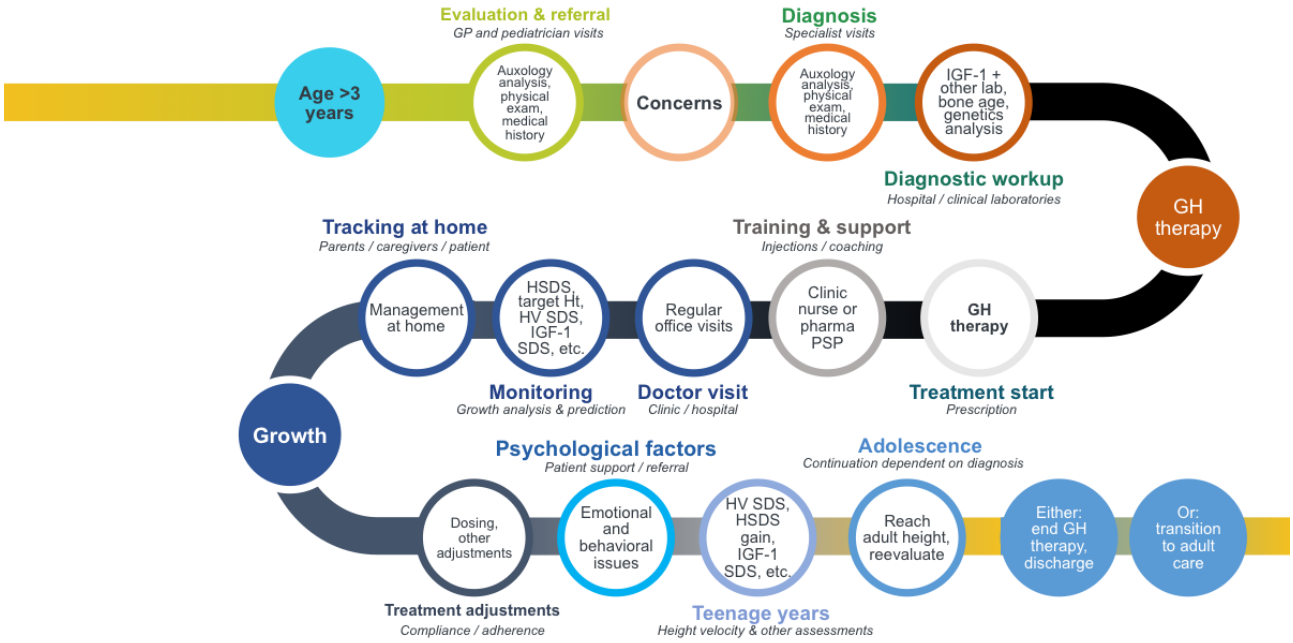
The case of growth hormone therapy is somewhat different from that of diabetes because patients do not require intensive or large-scale lifestyle modifications. Because growth hormone therapy is administered by daily injection, the most important self-management behavior is adherence to medication, which affects long-term outcome. Growth hormone therapy for children with short stature must be administered consistently and with a high level of adherence over many years for an effective gain in height and optimal adult height [17]. Therapy is required from early childhood to adolescence; therefore, support and involvement of the patient's family and caregivers are crucial to achieving the best outcomes. Thus, eHealth tools to support growth hormone therapy must be developed with a holistic approach that involves the family and must also take into consideration physiological, psychological, and developmental changes as children grow, with a view to using these tools over many years [18,19]. Additionally, some of the conditions that are supported with growth hormone therapy are classified as rare diseases (incidence <1 in 2500 [20]), posing extra challenges. Physicians' knowledge may be limited, which can often lead to delayed referral and diagnosis, and the small number of patients creates difficulties in collection of data to generate large and meaningful data sets unless performed on an international scale. The collection of adherence data is of crucial importance [21] to optimizing adherence and improving adult height [22,23].

An optimal adult height is attained through the early recognition and understanding of the different causes of growth failure, thus leading to early referral, diagnosis, and growth hormone therapy

XSL•FO
RenderX

initiation. As mapped out in Figure 1, outcomes can be improved at all stages of the patient journey through comprehensive monitoring of growth, education, and clinical support to achieve a high level of therapy adherence. We surmised that eHealth tools could play a significant role in the management of growth disorders and growth hormone therapy through additional and comprehensive support and monitoring.

Figure 1. Patient journey for children with growth failure and those who receive growth hormone therapy. GH: growth hormone; GP: general practitioner; HSDS: height standard deviation score; Ht: height; HV: height velocity; IGF-1: insulin-like growth factor-1; PSP: pediatric specialist physician; SDS: standard deviation score.



Objective

We aimed to develop a framework for using eHealth tools in the identification, investigation and diagnosis of growth failure and the management of growth hormone therapy in pediatric patients. Development of this framework was carried out as a multistep process, guided by the principles of participatory research involving key informants across a series of workshops and virtual activities. Similar approaches have been widely applied in medical informatics to identify implementation challenges of digital health tools [24], including the application of the Delphi method [25] in other endocrine disease areas [26] and eHealth for pediatrics [27].

Methods

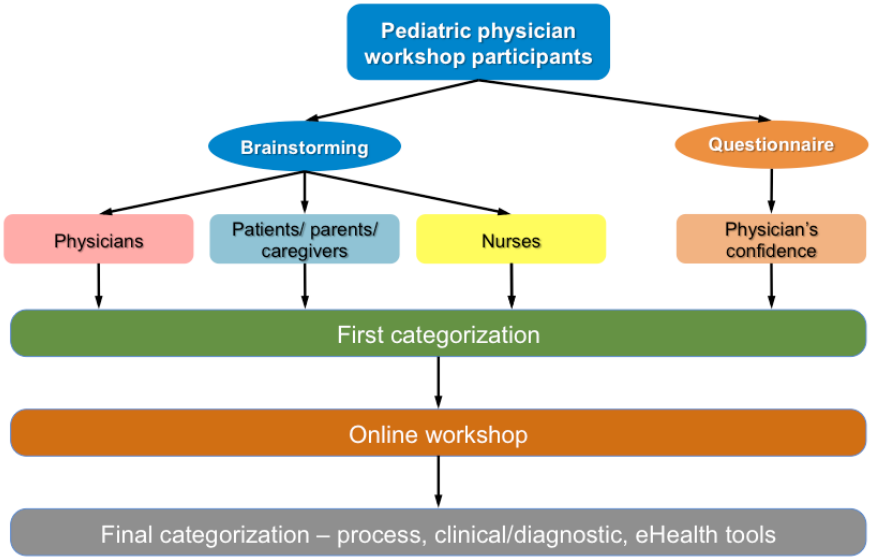
Initially, we conducted 3 online webinars with presentations and interactive discussions around key concepts and issues of eHealth, including data science and use in pediatric patients with growth failure and growth hormone therapy. Based on active involvement in this initial step, a group of 12 participants were subsequently invited to more detailed discussions. This report describes the approach that was adopted for these online detailed discussion workshops and is outlined (Figure 2) and explained in more detail in the following sections.

The 12 participants were pediatric endocrinologists who have experience in the treatment of growth disorders. Convenience sampling was used to select participants on the basis of previous participation, interest or experience in eHealth activities, and patients' use of connected injection devices for growth hormone therapy. Sampling also considered participants on the basis of providing worldwide geographical representation, with participants from Argentina, Brazil, Finland, France, Germany, Malaysia, Sweden, Spain, Taiwan, and the United Kingdom, encompassing a wide base of health care and socioeconomic conditions. The participants completed preparation activities that included reading relevant publications [28-30], brainstorming activities (see Multimedia Appendix 1), and completing an adapted version of the eHealth literacy scale (eHEALS) survey [31] related to digital health literacy (see Multimedia Appendix 2).

The brainstorming activities solicited opinions of the participants on opportunities, challenges, and established or potential eHealth tools relating to 3 stakeholder groups: (1) pediatric endocrinologists, (2) patients and their parents/caregivers, and (3) nurses. Simultaneously, the pediatric endocrinologists completed a survey about their perception of eHealth tools in aiding clinical decisions and their confidence in use of such tools, with results quantified as proportions of respondents agreeing or disagreeing with the statements on a 5-point scale. The brainstorming activities and surveys were designed using the SurveyMonkey tool (SVMK Inc).

XSL•FO
RenderX

Figure 2. Workshop participation methodological steps. Each step was carried out twice, first in relation to referral and diagnosis stages and second in relation to all stages from growth hormone therapy initiation to completion.



For the first online workshop, the physicians completed the brainstorming exercise and eHealth literacy questionnaire with respect to the first part of the patient journey (Figure 1), through referral and diagnosis. For the second online meeting, the exercise and questionnaire were completed with respect to the subsequent stages of the patient journey from start of growth hormone therapy, through monitoring, tracking, and assessment of outcomes, to either growth hormone discontinuation or into the transition phase from pediatric to adult care for patients still requiring growth hormone therapy.

In each case, the brainstorming responses were initially categorized according to common themes, which were then discussed in the online workshops under the moderation of PD who was the key facilitator and with additional availability of textual responses. The brainstorming responses for each of the 3 categories of opportunities, challenges, and eHealth tools were finally categorized as relating to process, clinical pathway (including diagnostics for the first meeting), or education. These 3 main categories reflected important aspects for the implementation of eHealth solutions in clinical practice and formed the basis for the framework for best practices: (1) process—responses that related to everyday tasks that aided the distance and time logistics required by the specified people for delivery of health care; (2) clinical—responses relating to the medical condition, including those relating to establishing diagnosis; and (3) education—responses relating to digital technologies that can enhance learning and understanding of the condition, and therapy for either health care specialists or the patients and families.

Results

Survey Responses

A total of 815 responses were received from the 2 brainstorming exercises, comprising 419 in relation to referral and diagnosis stages and 396 relating to growth hormone therapy stages. In each case, the responses were grouped and categorized; Figure 3 shows the grouped responses relating to physicians for the referral and diagnosis stages, while Figure 4 shows the growth hormone therapy stages (responses relating to patients/parents/caregivers can be found in Multimedia Appendix 3 for referral/diagnosis and Multimedia Appendix 4 for growth hormone therapy, and responses relating to nurses in Multimedia Appendix 5 for referral/diagnosis and Multimedia Appendix 6 for growth hormone therapy). In the following subsections we describe the responses provided, assigned to each group in turn.

XSL•FO
RenderX

Figure 3. Physician viewpoints regarding referral and diagnosis stages of pediatric patients with growth failure. Where no comments box is shown, clinicians did not provide any opinions relating to that category. HCP: health care provider.

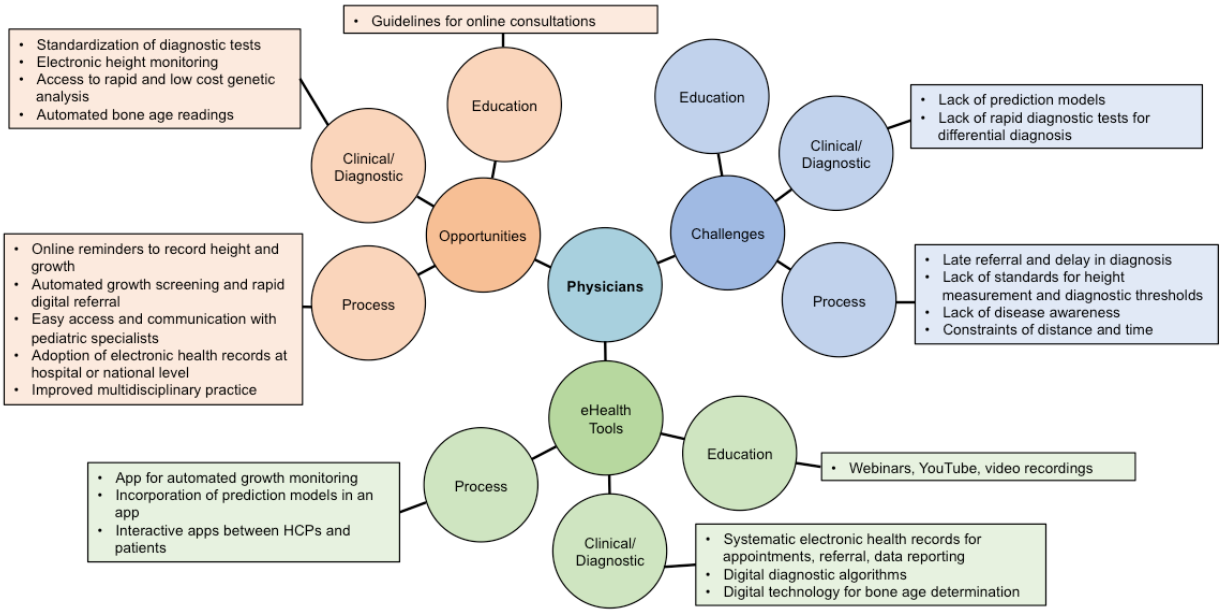
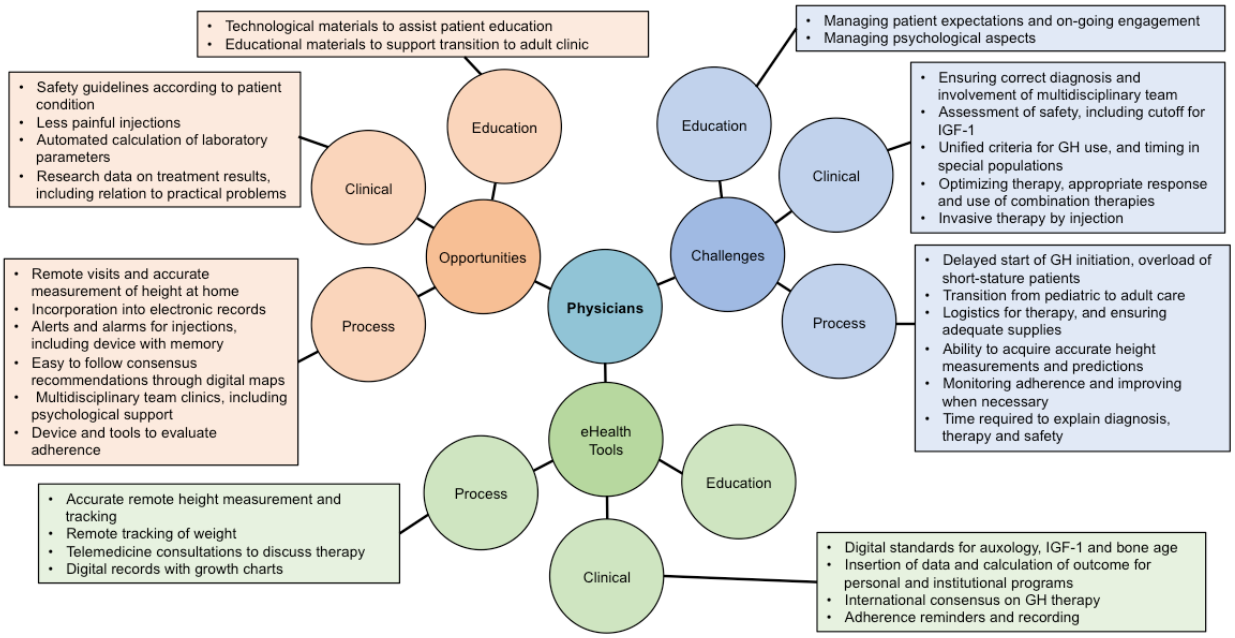


Figure 4. Physician viewpoints regarding growth hormone therapy initiation, monitoring, and transition. Where no comments box is shown, clinicians did not provide any opinions relating to that category. GH: growth hormone; IGF-1: insulin-like growth factor-1.



Brainstorming Responses Relating to Physicians

Growth Monitoring and Screening

For the referral and diagnosis stages, the perceived opportunities for physicians largely concerned process, particularly regarding digital reminders to record growth. It was noted that this could be done by primary care personnel, such as in Finland where children's height is routinely monitored and data immediately entered and integrated into electronic health records. Machine learning or artificial intelligence algorithms could be used to identify faltering growth and be followed by notification of health care professionals of this issue. Although such a system may overestimate referrals, it could support earlier diagnosis of conditions such as Turner syndrome and *SHOX* deficiency; it was noted that physicians may lack knowledge of these types of conditions, and pediatricians may not readily associate signs or symptoms with a diagnosis, resulting in late referral and therapy delay. Algorithms that have been developed to follow through symptomatic characteristics to identify the cause of growth failure could be incorporated into eHealth tools to aid physicians through the early stages of the referral and diagnostic process. Streamlining the screening, referral, and diagnostic process may expedite subsequent genetic testing, particularly for rare conditions associated with growth failure, thus identifying patients eligible for growth hormone therapy at an earlier stage.

XSL•FO
RenderX

Automated Height Measurements

At each stage of the patient journey, during referral, diagnosis, and monitoring of response to growth hormone therapy, the ability to accurately measure height electronically was cited as a significant area of development that could overcome the labor-intensive challenge of manually measuring height and inputting this into the patient record. There is potential for height and height velocity to be determined using electronic devices, such as mobile phones or tablet computers, which could be used by health professionals and parents or caregivers. Development of digital techniques for height measurement outside of the clinic would be very helpful in countries with remote rural communities where primary care is lacking, and data could be incorporated directly by linking to electronic health records. A tool that could be used by patients or caregivers to measure height electronically and be linked to artificial intelligence-based technology could increase the frequency of height measurements and provide a signal to health care providers about faltering growth to support earlier intervention.

Growth Hormone Response Prediction Models

The response to growth hormone therapy can be determined using validated prediction models that estimate potential adult height by incorporating parameters related to growth. These parameters include bone age, which is a determination of skeletal physiological maturity relative to chronological age through use of hand radiographs that can be evaluated with an automated digital method. There are multiple prediction models and a current lack of consensus regarding which of these is the gold standard. Clinicians wished to see a more unified international approach to growth prediction, which could subsequently lead to the incorporation of a model into an eHealth platform to aid clinicians in determining adult height.

Incorporation of prediction models into apps was considered a challenge, particularly given the lack of consensus on which model to use. It was, however, considered to be an important tool to allow identification of those patients who would most benefit from growth hormone therapy. Moreover, by using a digital device to measure height and then linking this to an electronic patient record, adult height prediction could be provided in a more seamless way through the integration of a prediction model into the process.

Communication, Logistics, and Engagement

Remote Delivery of Care and Communication

Communication between primary care and pediatric endocrine specialists, and others in multidisciplinary teams, was identified as an opportunity for eHealth tools to improve service provision. Multidisciplinary teams require good videoconferencing facilities; current platforms require improvements, such as increased bandwidth, to facilitate greater use of telemedicine. Guidelines for physicians that augment their ability to carry out online consultations are required, and education facilities for this could be provided through online workshops and video recordings.

One of the biggest challenges mentioned was constraints of time and distance when patients have to attend a hospital or clinic in order for assessments to be made. This may involve considerable travel over long distances (particularly for those in rural settings) and a large amount of time for both patients and health care personnel to ensure that appointments can be initiated and kept. Thus, the refinement of telecommunication systems will help to overcome this barrier.

Participants proposed that electronic health records could be used for automating appointments, referrals, and data reporting. As consultations improve efficiency, more information could be made available for each patient, giving interoperability greater importance in order that data entry does not require additional time. Overall, eHealth tools need to be efficient to use if they are to gain acceptance.

Patient Engagement

Digital technologies that can seamlessly record adherence with the therapy regimen allow of evaluation and timely intervention if adherence decreases over time. Adding adherence data to growth prediction models was considered a potential aid to the assessment of patients that would benefit from therapy and could be used as an additional variable in prediction of adult height.

Interactive games that motivate and engage the patient may also increase the long-term adherence to growth hormone therapy, and their development requires involvement of families, nurses, and physicians as key stakeholders.

Psychological Well-Being

One challenge for physicians is dealing with the psychological impact of growth hormone therapy. As patients move through different stages of their therapy journey, with evolving physical and emotional maturity, variations in therapy adherence are expected. Digital tools to evaluate the psychological status of patients could enable the identification and assessment of anxiety related to the condition and long-term therapy. Participants suggested that therapy-related psychology could be assessed using short online questionnaires, with standardization across different clinics. In addition, as online and digital tools are developed to deliver psychological interventions, these could be used to support patient care, thus reinforcing therapy adherence.

Evaluation of psychological status is particularly important at the end of pediatric care when patients who remain growth hormone–deficient transition to adult care in order to continue growth hormone therapy to optimize benefits of therapy beyond growth. Digital tools can be used to explain the pathway changes and specific needs to children and young adults, and can also provide feedback to refine care pathways.

Safety Assessments

Medical devices should also enable and encourage safety, with mechanisms built into the reporting platforms to individualize according to the specific diagnosis, therapy device, and condition of the patient. The reporting mechanism for safety concerns needs to be carefully considered to determine who should receive and respond to reports. Physicians could also be alerted to any problems in order to differentiate between

XSL•FO
RenderX

background symptoms of the condition and adverse events potentially related to therapy.

Brainstorming Exercise Responses From the Patient/Parent/Caregiver Perspective

Patient Education

Physician perspectives on opportunities and eHealth tools relating to patients, parents, or caregivers predominantly focused on education, largely concerning patient awareness, access to trusted information, and increased remote training. Videoconferencing tools are already being used for this to some extent, but more online tools involving apps, websites, medical portals, and search resources are needed to provide information that is accurate, trustworthy (ie, endorsed or produced by reputable sources with knowledge in the area), and easily understood. These tools could also be used to increase access to and interaction with health care personnel; however, access may need to be limited to avoid out-of-hours timing, but this could be mitigated by use of technologies such as virtual assistants or “chatbots ” to provide automated answers.

Patient-Reported Outcomes

Digital techniques to accurately measure height at home would enable growth to be monitored without the need for clinical review. These methods could also detect growth deceleration, with integration into electronic health records and prompting alerts sent to patients, parents, or caregivers. Therapy-related anxiety and psychological problems could be aided through games to generate interest in diagnosis and therapy, and to reduce needle anxiety. Appropriate tools could be developed to enhance psychological care, although it was noted that language barriers and digital access problems would need to be overcome to avoid digital exclusion and polarization of patients. Behavioral management may be enhanced through educational tools with a holistic approach, incorporating factors such as good nutrition and exercise to improve metabolic and physical outcomes beyond growth.

Brainstorming Exercise Responses From the Clinic Nurse Perspective

Responses in regard to nurses, similar to those from physicians, were more often related to process, including support for adherence, and on the medical devices used for therapy delivery. These patient-support tasks are frequently the responsibility of specialized nurses. Difficulties were noted in the delivery of roles, including patient testing, education. and ongoing support, which could be made easier by improved digitalization of and access to records, telemedicine technologies, and remote monitoring; eHealth tools could relieve nurses of some duties, yielding more time for personal interaction with the patient.

Responses also indicated challenges in the receipt and delivery of training and education to nurses around the management of growth failure and therapy, with limited opportunities for learning. In some counties, clinics do not have specialist nurses, and responsibilities may be inadequately defined, with no clear career paths. In such situations, online interactive learning resources could support training.

During growth hormone therapy, virtual reality or augmented reality immersive experiences could help with training and distraction during injections and blood sampling. During transition from pediatric to adult care, digital tools can assist nurses in providing psychological support and education for patients.

eHealth Literacy Survey

The questionnaire results indicated that the majority of the participants felt that eHealth tools were useful, but some were uncertain about evidence for their usefulness. There was also some uncertainty about what eHealth tools are currently available and how to use them. Most felt that digital tools are important and that they had the skills to use them, but many lacked confidence in their use largely due to the perceived lack of evidence. The respondents indicated that more data on the effectiveness of eHealth tools need to be presented for mainstream use in clinical practice, and interoperability would be required for routine clinical use.

Pediatric Growth Disorders and Growth Hormone Therapy Management Framework

From the workshop and questionnaire information, we devised a framework for developing eHealth tools for growth disorders and growth hormone therapy, as shown in Figure 5. The framework was devised to support integration of new eHealth tools into health care systems and assist health care professionals, patients, and families throughout the patient journey.

XSL•FO
RenderX

Figure 5. Framework for managing pediatric growth disorders and growth hormone therapy mapped to the patient journey. AR/VR: augmented reality/virtual reality; HCP: healthcare provider.



Discussion

Principal Findings

We are not aware of any similar studies and, to the best of our knowledge, this is the first time global experts in pediatric endocrinology have come together to consider possible eHealth solutions for the care of children and young people with growth disorders. Our work has led to greater understanding of the acceptability, confidence, and knowledge of the use of eHealth tools in growth disorders, and has opened up digital health care options to improve current clinical management.

Several key themes arose, which were aligned across the themed stakeholder groups considered: physicians, patients/parents/caregivers, and nurses. It was agreed that communication through established and novel technologies are likely to provide significant opportunities to improve interactions between clinicians across the health sector and to facilitate diagnosis, referral, and patient management. Importantly, telecommunication has the potential to improve management of patients who may be in isolated communities or rural areas remote to the central health care provider and to enable parents and caregivers to rapidly communicate with health care professionals when problems arise [32]. Connected digital systems will reduce travel time for families and reduce time of missed education, in turn supporting cost and efficiency savings for families and health services [33]. Integration of these systems into electronic patient records would optimize this process. Additionally, physicians suggested that parents and caregivers would value electronic processes to help them manage appointments, that administration tasks should be automated, and that improved interoperability between systems and within and across centers would allow sharing and integration of vital patient information.

Rapid development of technology could also facilitate the earlier identification of growth disorders through electronic screening programs, with the proviso that the input of height data is accurate. Thus, the development of a mobile phone or tablet-based tool that can accurately measure height in the home environment and that could link with a centralized electronic system may accelerate the identification and referral of patients with short stature and poor growth, reducing the prevalence of late diagnosis of rare disorders of growth where other comorbidities may exist [34-36]. However, a reliable electronic system for early identification of growth disorders is complex and will depend on the integration of systems that monitor height at home and in the clinical environment, with a high sensitivity and specificity [37].

Once a diagnosis of growth failure has been established, an accurate and validated digital tool to monitor height will provide a means of assessing response to growth hormone therapy. Response to growth hormone therapy and improved adherence are both predictors of adult height [23,38,39]. Fundamental to this is the use of artificial intelligence to understand and analyze patient growth data, which can be fed into an electronic health record. When patients become established on growth hormone, clinicians would value a unified means of predicting adult height, encompassing the diverse growth prediction models currently available [40-43]; identifying the impact of growth hormone through prediction models on digital platforms that link to or are embedded in patient records can aid optimization of therapy.

Education was identified as a key area to which eHealth tools could add significant value. Physicians, nurses, and families (including the patients) require relevant and trusted information delivered in an engaging, and developmentally and professionally appropriate way to support the rapid acquisition in knowledge [44,45]. Recent advances in medical education for patients and clinicians have focused on using immersive virtual and augmented reality (AR/VR), app-based interactive training, virtual assistants to answer questions in real-time, 3D modeling, and online real-time education [8,46,47]. In particular, virtual assistants or chatbots have been identified as a technology suitable for supporting patient groups with focused information available in real time [48-50]. One of the challenges in using novel technology-based education is ensuring that hard-to-reach and vulnerable groups have appropriate access to and understanding of these tools [51,52].

Participants also suggested that eHealth tools could be developed to support adherence to growth hormone therapy through the identification of therapy-related psychological and medical issues, potentially via access to data on patient-reported outcome measures [53] or ecological momentary assessments [54]. Given

XSL•FO
RenderX

the psychological and physiological changes that take place during growth and puberty, these tools will need to be versatile and responsive, supporting patients and their families during the transition from pediatric to adult health care settings [55]; furthermore, to mitigate the risk of future health complications, such tools need to reiterate education and prevent disengagement from health care [15,16,56]. Importantly, eHealth tools were also cited as an option to address some of the psychological issues identified, with digital platforms providing online psychological support; technologies including VR or AR could provide a means of reducing anxiety during painful procedures such as growth hormone injections, similar to the application of VR and AR as distraction tools in other conditions [8,46,56,57].

Although there was clear support for eHealth solutions to address challenges in the diagnosis and management of patients with growth disorders, results from the e-literacy questionnaire identified a need to ensure that the development and implementation of these tools is supported by a well-established evidence base and that integration within a clinical or home-based setting is supported by education and training.

The workshops and questionnaires have provided information that has led us to devise a framework to support the development of eHealth tools for growth disorders, which may also be relevant to other long-term conditions. The framework (Figure 5) has been devised as an eHealth overlay for the patient journey from identification of a problem, to diagnosis, implementation of therapy, and subsequent monitoring and support. The framework maps directly onto the patient pathway and thus will provide the relevant context for those developing novel technologies. It will support appropriate service and health care integration of new eHealth tools for health care professionals, patients, and families, which in turn will provide reassurance that technology development has appropriately addressed unmet needs for growth disorders.

Limitations

Participants in this study were chosen for their experience in childhood growth disorders and active interest and involvement in use of eHealth tools. However, participation was limited to physicians, albeit with significant experience in this field of medicine; thus, the views relating to patients, parents, caregivers, and nurses may be skewed by personal reflections and opinions, but responses were perhaps more directed and uniform. The relatively small number that provided responses and opinions may not be wholly representative of the pediatric endocrine community although they represented a wide geographical distribution, allowing us to consider the variation in health care systems in different countries. Due to logistic reasons, face-to-face meetings were not ideal, and we decided to conduct the workshops online, which might have restricted some activities that could have been better conducted through a face-to-face format. Although all participants were highly experienced pediatric endocrinologists, their differing backgrounds meant that there was an inherent variation in cultural perspectives, health care system setups, funding arrangements, and governmental priorities; however, this broad range of experiences could be viewed as reflective of real-world scenarios. Despite all participants having a strong interest in the use of digital techniques, eHealth tools may sometimes be seen to be purely theoretical, as they encompass technologies that could potentially be devised but are not yet established. The integration and adoption of our proposed eHealth framework into clinical practice remains to be assessed and rigorously tested in future studies.

Conclusions

By bringing together established and experienced pediatric endocrinologists from across the world, we have developed a framework for the implementation and integration of eHealth tools to support the identification and ongoing referral of children with growth disorders, and the subsequent management and monitoring of growth hormone therapy. Integration of eHealth tools in the management of patients with growth disorders will support height screening and regular height monitoring; improve communication; facilitate the delivery of and access to training and education for health care providers, patients, parents, and caregivers; improve therapy adherence; enable the earlier identification of problems; and support patients and families during their journey. Digitalization can enhance the collection of data from multiple sources to support research in low prevalence diseases, such as those treated with growth hormone, to build on best practice by providing an additional incentive for research collaboration using eHealth tools. The use of eHealth tools in the patient pathway has the potential to increase efficiency in care delivery while introducing cost benefits. The involvement of both service providers and service users is of crucial importance for the successful design and implementation of eHealth solutions. The credibility of these eHealth tools will be dependent on appropriate clinical evaluation of established and emerging digital tools. Moreover, training health care professionals and patients and their families in the use of eHealth tools while ensuring successful service integration and ongoing support will prevent the misuse or rejection of eHealth tools and unintended negative health consequences.

We conclude that novel eHealth tools have considerable potential for supporting patients and their families, and for health care providers in the detection and management of growth disorders. We envisage that our work may provide a blueprint for other chronic conditions that could benefit from eHealth tools in supporting long-term adherence.

Acknowledgments

This research is supported by Merck KGaA (Darmstadt, Germany). PD is also in receipt of research support for this study from the National Institute of Health Research (NIHR) Children and Young People MedTech Co-operative; the views expressed are those of the author and not necessarily those of the NHS (National Health Service) the NIHR, or of the Department of Health. The authors would like to thank Dr Peter Bates (Cambridge Medical Writing Services, United Kingdom) for medical writing

XSL•FO
RenderX

assistance, which was funded by Merck KGaA, Darmstadt, Germany. We would also like to thank David Dixon (Merck KGaA, Darmstadt, Germany) for his involvement in the study and helpful review of the manuscript.

Conflicts of Interest

EK is an employee of Merck KGaA, Darmstadt, Germany. LF-L is chief scientific officer at Adhera Health Inc, Palo Alto, CA, USA. MB is an employee of Tiba Medical Inc, Beaverton, OR, USA. TR provided paid advisory and consultancy services for Merck KGaA Darmstadt. The other authors have no conflicts to declare.

Multimedia Appendix 1
Online brainstorming assignment. Shown for referral and diagnosis stages and subsequently repeated for growth hormone therapy stages.
[DOCX File , 2127 KB - jmir_v23i5e27446_app1.docx]

Multimedia Appendix 2
eHealth survey. Shown for referral and diagnosis stages and subsequently repeated for growth hormone therapy stages.
[DOCX File , 686 KB - jmir_v23i5e27446_app2.docx]

Multimedia Appendix 3
Viewpoints on patient and parent/caregiver perspectives for referral and diagnosis stages.
[PPTX File , 81 KB - jmir_v23i5e27446_app3.pptx]

Multimedia Appendix 4
Viewpoints on patient and parent/caregiver perspectives for growth hormone therapy initiation, monitoring and transition stages.
[PPTX File , 86 KB - jmir_v23i5e27446_app4.pptx]

Multimedia Appendix 5
Viewpoints on nurse perspectives for referral and diagnosis stages.
[PPTX File , 84 KB - jmir_v23i5e27446_app5.pptx]

Multimedia Appendix 6
Viewpoints on nurse perspectives for growth hormone therapy initiation, monitoring, and transition stages.
[PPTX File , 83 KB - jmir_v23i5e27446_app6.pptx]

References

1. WHO Guideline: recommendations on digital interventions for health system strengthening. World Health Organization. Geneva; 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2021-05-06]
2. Sayeed R, Gottlieb D, Mandl KD. SMART Markers: collecting patient-generated health data as a standardized property of health information technology. NPJ Digit Med 2020;3:9. [doi: 10.1038/s41746-020-0218-6] [Medline: 31993507]
3. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state pf telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
4. Baltaxe E, Czypionka T, Kraus M, Reiss M, Askildsen JE, Grenkovic R, et al. Digital health transformation of integrated care in Europe: overarching analysis of 17 integrated care programs. J Med Internet Res 2019 Sep 26;21(9):e14956. [doi: 10.2196/14956]
5. Van Velthoven MH, Cordon C. Sustainable adoption of digital health innovations: perspectives from a stakeholder workshop. J Med Internet Res 2019 Mar 25;21(3):e11922. [doi: 10.2196/11922]
6. Brown SJ, Lieberman DA, Gemeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Medical Informatics 2009 Jul 12;22(1):77-89. [doi: 10.3109/14639239709089835]
7. Ratri DM, Fitri Hamidah K, Puspitasari AD, Farid M. Video-based health education to support insulin therapy in diabetes mellitus patients. J Public Health Res 2020 Jul 03;9(2):1849. [doi: 10.4081/jphr.2020.1849]
8. Calle-Bustos AM, Juan MC, García-García I, Abad F. An augmented reality game to support therapeutic education for children with diabetes. PLoS One 2017;12(9):e0184645 [FREE Full text] [doi: 10.1371/journal.pone.0184645] [Medline: 28957355]
9. Al-Taee MA, Kapoor R, Garrett C, Choudhary P. Acceptability of robot assistant in management of type 1 diabetes in children. Diabetes Technol Ther 2016 Dec;18(9):551-554. [doi: 10.1089/dia.2015.0428] [Medline: 27479198]

XSL•FO
RenderX

10. Navarro R, Dunn JD, Lee PA, Owens GM, Rapaport R. Translating clinical guidelines into practice: the effective and appropriate use of human growth hormone. Am J Manag Care 2013 Nov;19(15 Suppl):s281-s289 [FREE Full text] [Medline: 24494610]
11. Grimberg A, DiVall SA, Polychronakos C, Allen DB, Cohen LE, Quintos JB, Drug Therapeutics Ethics Committee of the Pediatric Endocrine Society. Guidelines for growth hormone and insulin-like growth factor-I treatment in children and adolescents: growth hormone deficiency, idiopathic short stature, and primary insulin-like growth factor-I deficiency. Horm Res Paediatr 2016;86(6):361-397 [FREE Full text] [doi: 10.1159/000452150] [Medline: 27884013]
12. Yuan Y, Zhou B, Liu S, Wang Y, Wang K, Zhang Z, et al. Meta-analysis of metabolic changes in children with idiopathic growth hormone deficiency after recombinant human growth hormone replacement therapy. Endocrine 2021 Jan;71(1):35-46. [doi: 10.1007/s12020-020-02435-w] [Medline: 32740695]
13. Dunger D, Darendeliler F, Kandemir N, Harris M, Rabbani A, Kappelgaard A. What is the evidence for beneficial effects of growth hormone treatment beyond height in short children born small for gestational age? A review of published literature. J Pediatr Endocrinol Metab 2020 Jan 28;33(1):53-70. [doi: 10.1515/jpem-2019-0098] [Medline: 31860471]
14. Quitmann J, Bloemeke J, Dörr HG, Bullinger M, Witt S, Silva N. First-year predictors of health-related quality of life changes in short-statured children treated with human growth hormone. J Endocrinol Invest 2019 Sep;42(9):1067-1076. [doi: 10.1007/s40618-019-01027-4] [Medline: 30840207]
15. Huang JS, Terrones L, Tompane T, Dillon L, Pian M, Gottschalk M, et al. Preparing adolescents with chronic disease for transition to adult care: a technology program. Pediatrics 2014 Jun;133(6):e1639-e1646 [FREE Full text] [doi: 10.1542/peds.2013-2830] [Medline: 24843066]
16. Sbardella E, Pozza C, Isidori A, Grossman A. Endocrinology and adolescence: dealing with transition in young patients with pituitary disorders. Eur J Endocrinol 2019 Oct;181(4):R155-R171. [doi: 10.1530/eje-19-0298] [Medline: 31370006]
17. Ranke MB, Lindberg A, Mullis PE, Geffner ME, Tanaka T, Cutfield WS, et al. Towards optimal treatment with growth hormone in short children and adolescents: evidence and theses. Horm Res Paediatr 2013;79(2):51-67 [FREE Full text] [doi: 10.1159/000347121] [Medline: 23446062]
18. Collin J, Whitehead A, Walker J. Educating children and families about growth hormone deficiency and its management: part 2. Nurs Child Young People 2016 Mar;28(2):30-36. [doi: 10.7748/ncyp.28.2.30.s23] [Medline: 26954647]
19. Grimberg A, Cousounis P, Cucchiara AJ, Lipman TH, Ginsburg KR. Parental concerns influencing decisions to seek medical care for a child's short stature. Horm Res Paediatr 2015;84(5):338-348 [FREE Full text] [doi: 10.1159/000440804] [Medline: 26448482]
20. Priority medicines for Europe and the world 2013 update. World Health Organization. 2013. URL: https://www.who.int/medicines/areas/priority_medicines/Ch6_19Rare.pdf [accessed 2021-05-06]
21. Koledova E, Tornincasa V, van Dommelen P. Analysis of real-world data on growth hormone therapy adherence using a connected injection device. BMC Med Inform Decis Mak 2020 Jul 29;20(1):176 [FREE Full text] [doi: 10.1186/s12911-020-01183-1] [Medline: 32727461]
22. Mohseni S, Heydari Z, Qorbani M, Radfar M. Adherence to growth hormone therapy in children and its potential barriers. J Pediatr Endocrinol Metab 2018 Jan 26;31(1):13-20. [doi: 10.1515/jpem-2017-0157] [Medline: 29216008]
23. Acerini CL, Wac K, Bang P, Lehwalder D. Optimizing patient management and adherence for children receiving growth hormone. Front Endocrinol (Lausanne) 2017;8:313 [FREE Full text] [doi: 10.3389/fendo.2017.00313] [Medline: 29209274]
24. Lupton D. Digital health now and in the future: Findings from a participatory design stakeholder workshop. Digit Health 2017;3:2055207617740018 [FREE Full text] [doi: 10.1177/2055207617740018] [Medline: 29942616]
25. de Meyrick J. The Delphi method and health research. Health Education 2003 Feb;103(1):7-16. [doi: 10.1108/09654280310459112]
26. Gabarron E, Bradway M, Fernandez-Luque L, Chomutare T, Hansen AH, Wynn R, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 05;18(1):414 [FREE Full text] [doi: 10.1186/s12913-018-3178-7] [Medline: 29871675]
27. Wald JS, Haque SN, Rizk S, Webb JR, Brown S, Ebron S, et al. Enhancing health IT functionality for children: the 2015 children's EHR format. Pediatrics 2018 Mar 08;141(4):e20163894. [doi: 10.1542/peds.2016-3894]
28. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
29. Agarwal S, Lefevre AE, Labrique AB. A call to digital health practitioners: new guidelines can help improve the quality of digital health evidence. JMIR Mhealth Uhealth 2017 Oct 06;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]
30. Digital therapeutics: combining technology and evidence-based medicine to transform personalized patient care. Digital Therapeutics Alliance. 2018 Oct. URL: https://dtxalliance.org/wp-content/uploads/2018/09/DTA-Report_DTx-Industry-Foundations.pdf [accessed 2021-01-26]
31. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
32. Ellis MJ, Russell K. The potential of telemedicine to improve pediatric concussion care in rural and remote communities in Canada. Front Neurol 2019;10:840 [FREE Full text] [doi: 10.3389/fneur.2019.00840] [Medline: 31428043]

XSL•FO
RenderX

33. Notario PM, Gentile E, Amidon M, Angst D, Lefaiver C, Webster K. Home-based telemedicine for children with medical complexity. Telemed J E Health 2019 Nov;25(11):1123-1132 [FREE Full text] [doi: 10.1089/tmj.2018.0186] [Medline: 30628860]
34. Scherdel P, Dunkel L, van Dommelen P, Goulet O, Salaün J, Brauner R, et al. Growth monitoring as an early detection tool: a systematic review. The Lancet Diabetes & Endocrinology 2016 May;4(5):447-456. [doi: 10.1016/s2213-8587(15)00392-7]
35. Murray PG, Dattani MT, Clayton PE. Controversies in the diagnosis and management of growth hormone deficiency in childhood and adolescence. Arch Dis Child 2016 Jan;101(1):96-100. [doi: 10.1136/archdischild-2014-307228] [Medline: 26153506]
36. Collett-Solberg PF, Ambler G, Backeljauw PF, Bidlingmaier M, Biller BMK, Boguszewski MCS, et al. Diagnosis, genetics, and therapy of short stature in children: a Growth Hormone Research Society international perspective. Horm Res Paediatr 2019;92(1):1-14 [FREE Full text] [doi: 10.1159/000502231] [Medline: 31514194]
37. Stalman SE, Hellinga I, van Dommelen P, Hennekam RC, Saari A, Sankilampi U, et al. Application of the Dutch, Finnish and British screening guidelines in a cohort of children with growth failure. Horm Res Paediatr 2015;84(6):376-382. [doi: 10.1159/000440652] [Medline: 26448202]
38. Ross JL, Lee PA, Gut R, Germak J. Attaining genetic height potential: Analysis of height outcomes from the ANSWER Program in children treated with growth hormone over 5 years. Growth Horm IGF Res 2015 Dec;25(6):286-293. [doi: 10.1016/j.ghir.2015.08.006] [Medline: 26363846]
39. van Dommelen P, Koledova E, Wit JM. Effect of adherence to growth hormone treatment on 0-2 year catch-up growth in children with growth hormone deficiency. PLoS One 2018;13(10):e0206009 [FREE Full text] [doi: 10.1371/journal.pone.0206009] [Medline: 30356273]
40. Wit JM, Ranke MB, Albertsson-Wikland K, Carrascosa A, Rosenfeld RG, Van Buuren S, et al. Personalized approach to growth hormone treatment: clinical use of growth prediction models. Horm Res Paediatr 2013;79(5):257-270 [FREE Full text] [doi: 10.1159/000351025] [Medline: 23735882]
41. Wikland KA, Kriström B, Rosberg S, Svensson B, Nierop AFM. Validated multivariate models predicting the growth response to GH treatment in individual short children with a broad range in GH secretion capacities. Pediatr Res 2000 Oct;48(4):475-484. [doi: 10.1203/00006450-200010000-00010] [Medline: 11004238]
42. Loftus J, Lindberg A, Aydin F, Gomez R, Maghnie M, Rooman R, et al. Individualised growth response optimisation (iGRO) tool: an accessible and easy-to-use growth prediction system to enable treatment optimisation for children treated with growth hormone. J Pediatr Endocrinol Metab 2017 Oct 26;30(10):1019-1026. [doi: 10.1515/jpem-2017-0120] [Medline: 28902629]
43. Smyczyńska U, Smyczyńska J, Hilczer M, Stawerska R, Tadeusiewicz R, Lewiński A. Pre-treatment growth and IGF-I deficiency as main predictors of response to growth hormone therapy in neural models. Endocr Connect 2018 Jan;7(1):239-249. [doi: 10.1530/ec-17-0277]
44. McCall M, Spencer E, Owen H, Roberts N, Heneghan C. Characteristics and efficacy of digital health education: An overview of systematic reviews. Health Educ J 2018 Aug;77(5):497-514. [doi: 10.1177/0017896918762013]
45. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education in health professions: the need for overarching evidence synthesis. J Med Internet Res 2019 Feb 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
46. Han S, Park J, Choi SI, Kim JY, Lee H, Yoo H, et al. Effect of immersive virtual reality education before chest radiography on anxiety and distress among pediatric patients: a randomized clinical trial. JAMA Pediatr 2019 Nov 01;173(11):1026-1031 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.3000] [Medline: 31498380]
47. Domhardt M, Tiefengrabner M, Dinic R, Fötschl U, Oostingh GJ, Stütz T, et al. Training of carbohydrate estimation for people with diabetes using mobile augmented reality. J Diabetes Sci Technol 2015 May;9(3):516-524. [doi: 10.1177/1932296815578880] [Medline: 25883165]
48. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
49. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
50. Beaudry J, Consigli A, Clark C, Robinson KJ. Getting ready for adult health care: designing a chatbot to coach adolescents with special health needs through the transitions of care. J Pediatr Nurs 2019;49:85-91. [doi: 10.1016/j.pedn.2019.09.004] [Medline: 31644960]
51. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):S131-S133 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
52. Smith B, Magnani JW. New technologies, new disparities: The intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]

XSL•FO
RenderX

53. Bele S, Chugh A, Mohamed B, Teela L, Haverman L, Santana MJ. Patient-reported outcome measures in routine pediatric clinical care: a systematic review. Front Pediatr 2020;8:364 [FREE Full text] [doi: 10.3389/fped.2020.00364] [Medline: 32850521]
54. 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]
55. Coyne I, Prizeman G, Sheehan A, Malone H, While AE. An e-health intervention to support the transition of young people with long-term illnesses to adult healthcare services: design and early use. Patient Educ Couns 2016 Dec;99(9):1496-1504. [doi: 10.1016/j.pec.2016.06.005] [Medline: 27372524]
56. Chan E, Hovenden M, Ramage E, Ling N, Pham JH, Rahim A, et al. Virtual reality for pediatric needle procedural pain: two randomized clinical trials. J Pediatr 2019 Jun;209:160-167.e4. [doi: 10.1016/j.jpeds.2019.02.034] [Medline: 31047650]
57. Dimitri P. Child health technology: shaping the future of paediatrics and child health and improving NHS productivity. Arch Dis Child 2019 Feb;104(2):184-188. [doi: 10.1136/archdischild-2017-314309] [Medline: 30154177]

Abbreviations

AR/VR: augmented reality/virtual reality
eHEALS: eHealth literacy scale
NIH: National Health Service
NIHR: National Institute of Health Research
SHOX: short-stature homeobox-containing gene

Edited by R Kukafka, G Eysenbach; submitted 26.01.21; peer-reviewed by C Athanasopoulou, B Nievas Soriano; comments to author 18.02.21; revised version received 22.03.21; accepted 11.04.21; published 20.05.21.

Please cite as:
Dimitri P, Fernandez-Luque L, Banerjee I, Bergadá I, Calliari LE, Dahlgren J, de Arriba A, Lapatto R, Reinehr T, Senniappan S, Thomas-Teinturier C, Tsai MC, Anuar Zaini A, Bagha M, Koledova E
An eHealth Framework for Managing Pediatric Growth Disorders and Growth Hormone Therapy
J Med Internet Res 2021;23(5):e27446
URL: https://www.jmir.org/2021/5/e27446
doi:10.2196/27446
PMID:34014174

©Paul Dimitri, Luis Fernandez-Luque, Indraneel Banerjee, Ignacio Bergadá, Luis Eduardo Calliari, Jovanna Dahlgren, Antonio de Arriba, Risto Lapatto, Thomas Reinehr, Senthil Senniappan, Cécile Thomas-Teinturier, Meng-Che Tsai, Azriyanti Anuar Zaini, Merat Bagha, Ekaterina Koledova. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Bots and Misinformation Spread on Social Media: Implications for COVID-19

McKenzie Himelein-Wachowiak[1], BA; Salvatore Giorgi[1,2], MSc; Amanda Devoto[1], PhD; Muhammad Rahman[1], PhD; Lyle Ungar[2], PhD; H Andrew Schwartz[3], PhD; David H Epstein[1], PhD; Lorenzo Leggio[1], MD, PhD; Brenda Curtis[1], MSPH, PhD

[1]Intramural Research Program, National Institute on Drug Abuse, Baltimore, MD, United States

[2]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Computer Science, Stony Brook Unversity, Stony Brook, NY, United States

Corresponding Author:
Brenda Curtis, MSPH, PhD
Intramural Research Program
National Institute on Drug Abuse
251 Bayview Blvd
Suite 200
Baltimore, MD, 21224
United States
Phone: 1 443 740 2126
Email: brenda.curtis@nih.gov

Abstract

As of March 2021, the SARS-CoV-2 virus has been responsible for over 115 million cases of COVID-19 worldwide, resulting in over 2.5 million deaths. As the virus spread exponentially, so did its media coverage, resulting in a proliferation of conflicting information on social media platforms—a so-called “infodemic.” In this viewpoint, we survey past literature investigating the role of automated accounts, or “bots,” in spreading such misinformation, drawing connections to the COVID-19 pandemic. We also review strategies used by bots to spread (mis)information and examine the potential origins of bots. We conclude by conducting and presenting a secondary analysis of data sets of known bots in which we find that up to 66% of bots are discussing COVID-19. The proliferation of COVID-19 (mis)information by bots, coupled with human susceptibility to believing and sharing misinformation, may well impact the course of the pandemic.

(J Med Internet Res 2021;23(5):e26933) doi:10.2196/26933

KEYWORDS

COVID-19; coronavirus; social media; bots; infodemiology; infoveillance; social listening; infodemic; spambots; misinformation; disinformation; fake news; online communities; Twitter; public health

Introduction

Globally, 2020 has been characterized by COVID-19, the disease caused by the SARS-CoV-2 virus. As of March 2021, the COVID-19 pandemic has been responsible for over 115 million documented cases, resulting in over 2.5 million deaths. The United States accounts for 24.9% of the world’s COVID-19 cases, more than any other country [1].

As the virus spread across the United States, media coverage and information from online sources grew along with it [2]. Among Americans, 72% report using an online news source for COVID-19 information in the last week, with 47% reporting that the source was social media [3]. The number of research articles focusing on COVID-19 has also grown exponentially; more research articles about the disease were published in the first 4 months of the COVID-19 pandemic than throughout the entirety of the severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) pandemics combined [4]. Unfortunately, this breadth, and the speed with which information can travel, sets the stage for the rapid transmission of misinformation, conspiracy theories, and “fake news” about the pandemic [5]. One study found that 33% of people in the United States report having seen “a lot” or “a great deal” of false or misleading information about the virus on social media [3]. Dr Tedros Adhanom Ghebreyesus, the Director-General of the World Health Organization, referred to this accelerated flow

XSL•FO
RenderX

of information about COVID-19, much of it inaccurate, as an "infodemic" [6].

Though the pandemic is ongoing, evidence is emerging regarding COVID-19 misinformation on social media. Rumors have spread about the origin of the virus, potential treatments or protections, and the severity and prevalence of the disease. In one sample of tweets related to COVID-19, 24.8% of tweets included misinformation and 17.4% included unverifiable information [7]. The authors found no difference in engagement patterns with misinformation and verified information, suggesting that myths about the virus reach as many people on Twitter as truths. A similar study demonstrated that fully false claims about the virus propagated more rapidly and were more frequently liked than partially false claims. Tweets containing false claims also had less tentative language than valid claims [8].

This trend of misinformation emerging during times of humanitarian crises and propagating via social media platforms is not new. Previous research has documented the spread of misinformation, rumors, and conspiracies on social media in the aftermath of the 2010 Haiti earthquake [9], the 2012 Sandy Hook Elementary School shooting [10], Hurricane Sandy in 2012 [11], the 2013 Boston Marathon bombings [12,13], and the 2013 Ebola outbreak [14].

Misinformation can be spread directly by humans, as well as by automated online accounts, colloquially called "bots." Social bots, which pose as real (human) users on platforms such as Twitter, use behaviors like excessive posting, early and frequent retweeting of emerging news, and tagging or mentioning influential figures in the hope they will spread the content to their thousands of followers [15]. Bots have been found to disproportionately contribute to Twitter conversations on controversial political and public health matters, although there is less evidence they are biased toward one "side" of these issues [16-18].

This paper combines a scoping review with an unpublished secondary analysis, similar in style to Leggio et al [19] and Zhu et al [20]. We begin with a high-level survey of the current bot literature: how bots are defined, what technical features distinguish bots, and the detection of bots using machine learning methods. We also examine how bots spread information, including misinformation, and explore the potential consequences with respect to the COVID-19 pandemic. Finally, we analyze and present the extent to which known bots are publishing COVID-19–related content.

What Are Bots?

Before addressing issues surrounding the spread of misinformation, we provide a definition of bots, describe their typical features, and explain how detection algorithms identify bots.

Definition and Identification

Bots, shorthand for "software robots," come in a large variety of forms. Bots are typically automated in some fashion, either fully automated or human-in-the-loop. There is a common conception that all bots spam or spread malware, but this is not the case. Some bots are benign, like the Twitter account @big_ben_clock, which impersonates the real Big Ben clock by tweeting the time every hour [21]. Others have even been used for social good, such as Botivist, which is a Twitter bot platform used for recruiting volunteers and donations [22]. Groups of bots can function in coordination with each other, forming what are called botnets [23]. One botnet of roughly 13,000 bot accounts was observed tweeting about Brexit, with most of these bot accounts disappearing from Twitter shortly after the vote [24]. Bots of all types are ubiquitous on social media and have been studied on Reddit [25,26], Facebook [27], YouTube [28], and Twitter [29], among other platforms.

Given their large variety, bots are often organized into subclasses, a selection of which we discuss here. Content polluters are one subclass; these are "accounts that disseminate malware and unsolicited content." Traditional spambots, another subclass, are "designed to be recognizable as bots" [30]. Social bots—a newer, more advanced type of bot [31-33]—use "a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior." There are also hybrid human-bot accounts (often called cyborgs) [34], which "exhibit human-like behavior and messages through loosely structured, generic, automated messages and from borrowed content copied from other sources" [35]. It is not always clear which category a bot may fall into (eg, if a given social bot is also a cyborg).

Various methods have been used to identify bots "in the wild," so as to build the data sets of known bots used to train bot-detection algorithms. One method, the "social honeypot" [36], mimics methods traditionally used by researchers to monitor hacker activity [37] and email harvesting [38]. Specifically, social honeypots are fake social media profiles set up with characteristics desirable to spammers, such as certain demographics, relationship statuses, and profile pictures [39]. When bots attempt to spam the honeypots (by linking malware-infested content or pushing product websites), researchers can easily identify them.

Technical Features of Bots

Overview

Features that distinguish bots from humans roughly fall into three categories: (1) network properties, such as hashtags and friend/follower connections, (2) account activity and temporal patterns, and (3) profile and tweet content. These feature categories have the advantage of being applicable across different social media platforms [27].

Network Properties

Networks based on friend/follower connections, hashtag use, retweets, and mentions have been used in a number of studies that seek to identify social bots [40-43], exploiting network homophily (ie, humans tend to follow other humans and bots tend to follow other bots). As bots become more sophisticated, network properties become less indicative of them; studies have found groups of bots that were able to build social networks that mimic those of humans [44].

XSL•FO
RenderX

Account Activity and Temporal Patterns

Patterns of content generation can be good markers of bots. Bots compose fewer original tweets than humans, but retweet others' tweets much more frequently, and have a shorter time interval between tweets [40]. Ferrara et al [31] found that humans are retweeted by others more than are bots, suggesting that bots may struggle to compose convincing or interesting tweets. However, many others have found this not to be the case [15,16,33]. Finally, humans typically modify their behavior during each online session; as the session progresses, the density of new tweets decreases. Bots do not engage in these "sessions" of social media usage, and accordingly do not modify their behavior [45].

Profile and Tweet Content

Profile metadata such as account age and username can be used to identify social bots. Ferrara et al [31] showed that bots have shorter account age (ie, the accounts were created more recently), as well as longer usernames. Automatic sentiment analysis of tweet content has also been studied as a means of distinguishing bots from humans. One study found humans expressed stronger positive sentiment than bots, and that humans more frequently "flip-flopped" in their sentiment [42].

Detection of Bots

Over the past decade, several teams have sought to develop algorithms that successfully identify bots online. Social media platforms use similar algorithms internally to remove accounts likely to be bots. These algorithms originated in early attempts to identify spam emails [46], social phishing [47], and other types of cybercrimes [37]. With the advent of online communities, cybercriminals turned their attention to these sites, eventually creating fake, automated accounts at scale [48]. Table 1 provides a summary of several prominent papers on bot identification. We note that the details of specific machine learning algorithms are beyond the scope of this paper and therefore are not included in this manuscript.

Table 1. Review of state-of-the-art detection of bots on Facebook and Twitter.

Type and reference	Platform	Number of accounts	Features			Model	Metric	Predictive accuracy
			N[a]	T[b]	C[c]			
Human judgment (manual annotation)								
Cresci et al (2017) [33]	Twitter	928				Manual annotation	F1-score	0.57
Automatic methods								
Ahmed and Abulaish (2013) [27]	Facebook and Twitter	320 (Facebook), 305 (Twitter)			✓	Naïve Bayes, decision trees, rule learners	Detection rate	0.96 (Facebook), 0.99 (Twitter)
Dickerson et al (2014) [42]	Twitter	897	✓	✓	✓	Gradient boosting	Area under the curve	0.73
Cresci et al (2017) [33]	Twitter	928		✓		Digital DNA sequences	F1-score	0.92
Varol et al (2017) [41]	Twitter	21,000	✓	✓	✓	Random forests	Area under the curve	0.95
Kudugunta and Ferrara (2018) [49]	Twitter	8386			✓	AdaBoost	Area under the curve	>0.99
Mazza et al (2019) [50]	Twitter	1000		✓		Long short-term memory networks	F1-score	0.87
Santia et al (2019) [51]	Facebook	1000			✓	Support vector machines, decision trees, Naïve Bayes	F1-score	0.72
Yang et al (2020) [52]	Twitter	137,520		✓	✓	Random forests	Area under the curve	0.60-0.99

[a]N: network properties.

[b]T: account activity and temporal patterns.

[c]C: profile and tweet content.

The first reference in Table 1 involved a manual annotation task in which raters were asked to label a Twitter account as human or bot. The fourth study listed in the table is the same as the first study; in this study, the same data set was evaluated by both human annotators and machine learning methods [33]. There was a large discrepancy in predictive accuracy (F1-score) between the two methods: 0.57 for the human annotators versus 0.92 for the automated method. Stated another way, human participants correctly identified social bots less than 25% of the time, though they were quite good at identifying genuine (human) accounts (92%) and traditional spambots (91%). These results suggest that social bots have a very different online presence from traditional spambots, or "content polluters"—and that this presence is convincingly human. Even if the human annotators are compared to the lowest scoring automated method (which we note is in a different domain and, thus, not directly comparable), the machine learning algorithm still provides a considerable boost in F1-score (0.57 versus 0.72).

XSL•FO
RenderX

There is no good way to compare all automated methods directly, as data sets are typically built in a single domain (ie, a single social media platform) and rapid advances in machine learning techniques prevent comparisons between models published even a few years apart. Furthermore, results suggest that models trained on highly curated bot data sets (eg, groups of accounts promoting certain hashtags or spamming a particular honeypot) may not perform well at detecting bots in other contexts. Yang et al [52] used a large number of publicly available bot data sets, training machine learning models on each set and testing them on those remaining. The result was a wide range of predictive accuracies across different bot data sets.

How Do Bots Amplify and Spread Misinformation?

We adopt the definition of misinformation used by Treen and colleagues: "misleading [or false] information that is created and spread, regardless of intent to deceive" [53]. For the purposes of this paper, we include fake news and false conspiracy theories under this umbrella term.

Many features of bots likely enable them to be "super-spreaders" of misinformation. Bots have been shown to retweet articles within seconds of their first being posted, contributing to the articles going viral [15]. Moreover, the authors of this study found that 33% of the top sharers of content from low-credibility sources were likely to be bots, significantly higher than the proportion of bots among top sharers of fact-checked content. Similarly, in a study of bots and "anti-vaxxer" tweets, Yuan et al [18] found that bots were "hyper-social," disproportionately contributing to content distribution. Bots also employ the strategy of mentioning influential users, such as @realDonaldTrump, in tweets linking to false or misleading articles, and are more likely to do so than their human counterparts [15]. The hope is that these users will share the article with their many followers, contributing to its spread and boosting its credibility. "Verified" (blue check) Twitter users, often celebrities, have been shown to both author and propagate COVID-19–related misinformation [54]. Interestingly, the frequency of false claims about the 2020 election dropped dramatically in the week after former president Donald Trump was removed from the platform [55].

In light of findings that humans are largely unable to distinguish social bots from genuine (human) accounts [33], it is likely that humans unknowingly contribute to the spread of misinformation as well. Accordingly, one study found that in regard to low-credibility content, humans retweet bots and other humans at the same rate [15]. Similarly, Vosoughi et al [56] found that "fake news" articles spread faster on Twitter than true news articles because humans, not bots, were more likely to retweet fake articles. Given human susceptibility to both automated accounts and "fake news," some have warned that intelligent social bots could be leveraged for mass deception or even "political overthrow" [57].

There is reason to believe that bots have already infiltrated political conversations online. Leading up to the 2016 presidential election in the United States, 20% of all political tweets originated from accounts that were likely to be bots [16]. While it did not specifically implicate bots, one study found that a majority of "fake" or extremely biased news articles relating to the 2016 election were shared by unverified accounts—that is, accounts that were not confirmed to be human [58]. There is also evidence that bots spread misinformation in the 2017 French presidential election, though ultimately the bot campaign was unsuccessful, in part because the human users who engaged with the bots were mostly foreigners with no say in the election outcome [59]. Bot strategies specifically relevant to political campaigns include "hashtag hijacking," in which bots adopt an opponent's hashtags in order to spam or otherwise undermine them, as well as flagging their opponent's legitimate content in the hopes it gets removed from the platform [60].

Where Do Bots Come From?

The origin of social bots is a challenging question to answer. Given the aforementioned concerns of political disruption by social bots, one may assume that foreign actors create social bots to interfere with political processes. Indeed, the Mueller report found evidence of Russian interference in the 2016 US election via social media platforms [61], and Twitter reports removing over 50,000 automated accounts with connections to Russia [62]. However, locating the origin of a social media account is difficult, as tweets from these accounts are rarely geotagged. Rheault and Musulan [63] proposed a methodology to identify clusters of foreign bots used during the 2019 Canadian election using uniform manifold approximation and projection combined with user-level document embeddings. Simply put, the authors constructed communities of users via linguistic similarities, and identified members significantly outside these communities as foreign bots.

Of note, studies have shown that a majority of social bots focusing on election-related content originate domestically [63]. Reasons for a candidate or their supporters to employ social bots may be relatively benign, such as boosting follower counts or sharing news stories, or they may involve smear campaigns [64].

While the ability to investigate the origin and motive of social bots is difficult, the means to create a social bot are fairly easy to access. Social bots are available for purchase on the dark web, and there are tens of thousands of codes for building social bots on free repositories like GitHub [65]. Of note, the top contributors of bot-development tools for mainstream social media sites are the United States, the United Kingdom, and Japan. The authors of this paper also note the intelligence and capabilities of these freely available bots may be overstated.

Are Bots Tweeting About COVID-19?

In light of the COVID-19 "infodemic" and findings that social bots have contributed to misinformation spread in critical times, we sought to assess the number of known Twitter bots producing COVID-19–related content. To this end, we gathered a number of publicly available bot data sets from the Bot Repository [66]. These data sets include both traditional spambots and social

XSL•FO
RenderX

bots that were first identified through a number of different methods (see the original papers for more details).

Using the open-source Python package TwitterMySQL [67], which interfaces with the Twitter application programming interface (API), we were able to pull all tweets from 2020 for each bot in the combined data set. Of note, Twitter's API limits access to tweets and account information available at the time of collection. Tweets and accounts that have been deleted or made private since originally appearing in one of the above papers are not made available, meaning we had less data than what was reported in the original papers. Our final data set consisted of 3.03 million tweets from 3953 bots, with an average of 768.9 (SD 1145.4) tweets per bot, spanning January 1, 2020, to August 21, 2020.

From these data, we pulled tweets using a set of 15 COVID-19–related keywords, which have previously been used to identify COVID-19 tweets in a study tracking mental health and psychiatric symptoms over time [68]. Sample keywords include #coronavirus, #covid19, and #socialdistancing. We then counted the number of accounts that mentioned these keywords in tweets since January 2020. Table 2 shows the percentage of bot accounts in each data set currently tweeting about COVID-19. Original sample size refers to the number of bots identified in this data set, while current sample size is the number of currently active bots (ie, tweeting in 2020). Between 53% (96/182) and 66% (515/780) of these bots are actively tweeting about COVID-19.

Table 2. Open-source data sets of bots discussing COVID-19[a].

Reference	Year	Original sample size, n	Current sample size, n	Bots discussing COVID-19, n (%)
Lee et al [36]	2011	22,223	2623	1427 (54)
Varol et al [41]	2017	826	292	164 (56)
Gilani et al [69]	2017	1130	780	515 (66)
Cresci et al [33]	2017	4912	77	48 (62)
Mazza et al [50]	2019	391	182	96 (53)

[a]Original sample size is the number of bot IDs publicly released on the Bot Repository, while current sample size is the number of active accounts tweeting in 2020. Percentage discussing COVID-19 is the percentage of bots with at least one tweet containing a COVID-19 keyword out of those active in 2020.

Implications for the COVID-19 Pandemic

Here we have shown that a majority of known bots are tweeting about COVID-19, a finding that corroborates similar studies [68,70]. Early in the pandemic, one study found that 45% of COVID-19–related tweets originate from bots [71], although Twitter has pushed back on this claim, citing false-positive detection algorithms [72]. Another study showed that COVID-19 misinformation on Twitter was more likely to come from unverified accounts—that is, accounts not confirmed to be human [7]. In an analysis of 43 million COVID-19–related tweets, bots were found to be pushing a number of conspiracy theories, such as QAnon, in addition to retweeting links from partisan news sites [73]. Headlines from these links often suggested that the virus was made in Wuhan laboratories or was a biological weapon.

One limitation of our study is that we did not investigate the validity of COVID-19–related claims endorsed by bots in our analyses. It may be that bots are largely retweeting mainstream news sources, as was the case in a recent study of bots using #COVID19 or #COVID-19 hashtags [68]. However, previous research has connected bots to the spread of misinformation in other public health domains, such as vaccines [30] and e-cigarettes [74], and unsubstantiated medical claims surrounding the use of marijuana [75].

Such misinformation can have detrimental consequences for the course of the COVID-19 pandemic. Examples of these real-world consequences include shortages of hydroxychloroquine (a drug that is crucial for treating lupus and malaria) due to increased demand from people who believe it will protect them from COVID-19 [76,77]. This drug has been promoted as a preventative against COVID-19 on social media, even though several randomized controlled trials have found it ineffective, [78,79], and the National Institutes of Health recently halted its own trial due to lack of effectiveness [80]. Moreover, belief in conspiracy theories about COVID-19 is associated with a decreased likelihood of engaging in protective measures such as frequent handwashing and social distancing, suggesting that misinformation may even contribute to the severity of the pandemic [81]. In addition, exposure to misinformation has been negatively correlated with intention to take a COVID-19 vaccine [82].

We are certainly not the first to express concern with viral misinformation; in May 2020, Twitter began labeling fake or misleading news related to COVID-19 in an effort to ensure the integrity of information shared on the platform [83]. Facebook introduced even more controls, such as organizing the most vetted articles at the top of the news feed, banning antimask groups, and sending antimisinformation messages to users who have shared fake news [84]. However, these measures are designed to target humans. In light of the numerous viral rumors relating to COVID-19 and the US response to the pandemic, we believe that bots likely contributed to their spread.

Major social media platforms like Twitter and Facebook do have methods to curtail suspected bots. In 2018, Twitter banned close to 70 million suspicious accounts in a matter of months [85]. Facebook banned 1.3 billion suspicious accounts in the third quarter of 2020. The platform estimates these accounts represent 5% of its worldwide monthly active users. The vast

XSL•FO
RenderX

majority of suspicious accounts were identified using automated detection methods, but 0.7% were first flagged by human users, suggesting that everyday Facebook users concerned about malicious activity on the platform can contribute to efforts to ban these accounts [86].

Mitigation of the harmful effects of social bots can also occur at the policy level. In 2018, California became the first and only state to pass a law requiring social bots to identify themselves as such [87]. In 2019, Senator Dianne Feinstein proposed a similar bill federally; the bill would allow the Federal Trade Commission to enforce bot transparency and would prohibit political candidates from incorporating social bots in their campaign strategy [88]. The United States Congress has brought top executives from Facebook, Twitter, and Google to testify before Congress about Russian influence on their platforms in advance of the 2016 election [89]. Scholars have interpreted these actions as a sign that the government wishes to maintain the right to regulate content on social media—a prospect that brings concerns of its own [90]. Presently, content problems on social media platforms are almost exclusively dealt with by the owners of those platforms, usually in response to user complaints, but in the coming years we may see an increase in government oversight on these platforms, fueling concerns about state-sponsored censorship [91,92]. More fundamentally, some have argued that, before any actionable policy or automatic interventions can be enabled, ambiguities in both bot definitions and jurisdiction and authority need to be addressed [90].

Even as citizens, social media platforms, and policy makers converge on the notion that bots and misinformation are urgent problems, the methods used to address the issue have had mixed results. When social media platforms crack down on bots and misinformation, either through automated techniques or manual content moderation, they run the risk of censoring online speech and further disenfranchising minority populations. Content promotion and moderation can lead to arbitrary policy decisions that may be inconsistent across or even within platforms [93]. In one example, Facebook ignited a controversy when their moderators flagged a breastfeeding photo as obscene, leading to a large number of protests on both sides of the debate [94]. Automated methods suffer from similar drawbacks, with multiple studies showing that biases in machine learning models can have unintended downstream consequences [95]. For example, algorithms designed to detect hate speech were more likely to label a post as “toxic” when it showed markers of African American English [96].

Finally, there is a continued arms race between bot-detection algorithms and bot creators [21,33]. As bots inevitably become more intelligent and convincingly human, the means for identifying them will have to become more precise. We observed that the majority of known bots in a sample of publicly available data sets are now tweeting about COVID-19. These bots, identified between 2011 and 2019, were discovered before the pandemic and were originally designed for non–COVID-19 purposes: promoting product hashtags, retweeting political candidates, and spreading links to malicious content. The COVID-19 pandemic will eventually end, but we have reason to believe social bots, perhaps even the same accounts, will latch on to future global issues. Additionally, we can expect bot generation techniques to advance, especially as deep learning methods continue to improve on tasks such as text or image generation [97,98]. Bot creators will continue to deploy such techniques, possibly fooling detection algorithms and humans alike. In the end, we should not expect current detection techniques, self-policing of social media platforms, or public officials alone to fully recognize, or adequately address, the current landscape of bots and misinformation.

Acknowledgments

This work was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Drug Abuse. MHW wrote the manuscript, with help from SG and AD. SG conceptualized the paper, with input from BC, HAS, and LU. SG performed the analyses on Twitter bots and created Tables 1 and 2. MHW, SG, AD, and MR all contributed to the literature review. LL and DHE provided crucial edits. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Coronavirus Resource Center. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). URL: https://coronavirus.jhu.edu/map.html [accessed 2020-12-21]
2. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
3. Nielsen RK, Fletcher R, Newman N, Brennen JS, Howard PN. Navigating the 'infodemic': How people in six countries access and rate news and information about coronavirus.: Reuters Institute; 2020 Apr. URL: http://www.fundacionindex.com/fi/wp-content/uploads/2020/04/Navigating-the-Coronavirus-Infodemic-FINAL.pdf [accessed 2021-05-09]
4. Valika TS, Maurrasse SE, Reichert L. A Second Pandemic? Perspective on Information Overload in the COVID-19 Era. Otolaryngol Head Neck Surg 2020 Nov;163(5):931-933. [doi: 10.1177/0194599820935850] [Medline: 32513072]
5. Ball P, Maxmen A. The Epic Battle Against Coronavirus Misinformation and Conspiracy Theories. Nature 2020 May;581(7809):371-374. [doi: 10.1038/d41586-020-01452-z] [Medline: 32461658]

XSL•FO
RenderX

6. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
7. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
8. Shahi G, Dirkson A, Majchrzak T. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021 Mar;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
9. Oh O, Kwon KH, Rao HR. An exploration of social media in extreme events: Rumor theory and Twitter during the Haiti earthquake 2010. 2010 Presented at: International Conference on Information Systems 2020; 2010; St. Louis, MO URL: https://aisel.aisnet.org/icis2010_submissions/231
10. Williamson E. How Alex Jones and Infowars helped a Florida man torment Sandy Hook families. The New York Times. 2019 Mar 29. URL: https://www.nytimes.com/2019/03/29/us/politics/alex-jones-infowars-sandy-hook.html [accessed 2021-05-09]
11. Wang B, Zhuang J. Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. Nat Hazards 2018 May 11;93(3):1145-1162 [FREE Full text] [doi: 10.1007/s11069-018-3344-6]
12. Gupta A, Lamba H, Kumaraguru P. $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter. 2013 Presented at: 2013 APWG eCrime Researchers Summit; September 17-18, 2013; San Francisco, CA p. 1-12 URL: https://doi.org/10.1109/eCRS.2013.6805772 [doi: 10.1109/ecrs.2013.6805772]
13. Starbird K, Maddock J, Orand M, Achterman P, Mason RM. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In: iConference 2014 Proceedings. 2014 Presented at: iConference 2014; March 4-7, 2014; Berlin, Germany p. 654-662. [doi: 10.9776/14308]
14. Jin F, Wang W, Zhao L, Dougherty E, Cao Y, Lu C, et al. Misinformation Propagation in the Age of Twitter. IEEE Annals of the History of Computing 2014 Dec;47(12):90-94. [doi: 10.1109/MC.2014.361]
15. Shao C, Ciampaglia GL, Varol O, Yang K, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018 Nov 20;9(1):4787 [FREE Full text] [doi: 10.1038/s41467-018-06930-7] [Medline: 30459415]
16. Bessi A, Ferrara E. Social bots distort the 2016 U.S. Presidential election online discussion. First Monday 2016 Nov 03;21(11):1 [FREE Full text] [doi: 10.5210/fm.v21i11.7090]
17. Badawy A, Ferrara E, Lerman K. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. 2018 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 28-31, 2018; Barcelona, Spain p. 258-265. [doi: 10.1109/asonam.2018.8508646]
18. Yuan X, Schuchard RJ, Crooks AT. Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter. Social Media + Society 2019 Sep 04;5(3):205630511986546 [FREE Full text] [doi: 10.1177/2056305119865465]
19. Leggio L, Garbutt JC, Addolorato G. Effectiveness and safety of baclofen in the treatment of alcohol dependent patients. CNS Neurol Disord Drug Targets 2010 Mar;9(1):33-44. [doi: 10.2174/187152710790966614] [Medline: 20201813]
20. Zhu Y, Guntuku SC, Lin W, Ghinea G, Redi JA. Measuring Individual Video QoE: A survey, and proposal for future directions using social media. ACM Trans Multimedia Comput Commun Appl 2018 May 22;14(2s):1-24. [doi: 10.1145/3183512]
21. Yang KC, Varol O, Davis CA, Ferrara E, Flammini A, Menczer F. Arming the public with artificial intelligence to counter social bots. Hum Behav & Emerg Tech 2019 Feb 06;1(1):48-61. [doi: 10.1002/hbe2.115]
22. Savage S, Monroy-Hernandez A, Hollerer T. Botivist: Calling volunteers to action using online bots. 2016 Presented at: 16th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 2, 2016; San Francisco, CA p. 813-822. [doi: 10.1145/2818048.2819985]
23. Abokhodair N, Yoo D, McDonald DW. Dissecting a Social Botnet: Growth, Content and Influence in Twitter. 2015 Presented at: 18th ACM Conference on Computer-Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, Canada p. 839-851. [doi: 10.1145/2675133.2675208]
24. Bastos MT, Mercea D. The Brexit Botnet and User-Generated Hyperpartisan News. Social Science Computer Review 2017 Oct 10;37(1):38-54. [doi: 10.1177/0894439317734157]
25. Jhaver S, Bruckman A, Gilbert E. Does Transparency in Moderation Really Matter?: User Behavior After Content Removal Explanations on Reddit. Proceedings of the ACM on Human-Computer Interaction 2019;3(CSCW):1-27. [doi: 10.1145/3359252]
26. Ma MC, Lalor JP. An Empirical Analysis of Human-Bot Interaction on Reddit. : Association for Computational Linguistics; 2020 Presented at: Sixth Workshop on Noisy User-generated Text (W-NUT 2020); November 2020; Virtual p. 101-106. [doi: 10.18653/v1/2020.wnut-1.14]
27. Ahmed F, Abulaish M. A generic statistical approach for spam detection in Online Social Networks. Computer Communications 2013 Jun;36(10-11):1120-1129. [doi: 10.1016/j.comcom.2013.04.004]

XSL•FO
RenderX

28. Hussain MN, Tokdemir S, Agarwal N, Al-Khateeb S. Analyzing disinformation and crowd manipulation tactics on YouTube. : IEEE; 2018 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 28-31, 2018; Barcelona, Spain p. 1092-1095. [doi: 10.1109/asonam.2018.8508766]
29. Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, et al. The DARPA Twitter Bot Challenge. Computer 2016 Jun;49(6):38-46. [doi: 10.1109/MC.2016.183] [Medline: 27295638]
30. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
31. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]
32. Zhang J, Zhang R, Zhang Y, Yan G. The Rise of Social Botnets: Attacks and Countermeasures. IEEE Trans Dependable and Secure Comput 2018 Nov 1;15(6):1068-1082. [doi: 10.1109/tdsc.2016.2641441]
33. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. 2017 Presented at: The 26th International Conference on World Wide Web Companion; April 3-7, 2017; Perth, Australia p. 963-972. [doi: 10.1145/3041021.3055135]
34. Chu Z, Gianvecchio S, Wang H, Jajodia S. Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? IEEE Trans Dependable and Secure Comput 2012 Nov;9(6):811-824. [doi: 10.1109/Tdsc.2012.75]
35. Clark EM, Williams JR, Jones CA, Galbraith RA, Danforth CM, Dodds PS. Sifting robotic from organic text: A natural language approach for detecting automation on Twitter. Journal of Computational Science 2016 Sep;16:1-7 [FREE Full text] [doi: 10.1016/j.jocs.2015.11.002]
36. Lee K, Eoff B, Caverlee J. Seven months with the devils: A long-term study of content polluters on Twitter. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media (ICWSM); July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2780/3296
37. Spitzner L. The Honeynet Project: Trapping the hackers. IEEE Secur Privacy 2003 Mar;1(2):15-23. [doi: 10.1109/msecp.2003.1193207]
38. Prince MB, Dahl BM, Holloway L, Keller AM, Langheinrich E. Understanding how spammers steal your email address: An analysis of the first six months of data from Project Honey Pot. 2005 Presented at: Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS); July 21-22, 2005; Stanford, CA URL: https://www.cse.scu.edu/~tschwarz/coen252_07/Resources/spammer.pdf
39. Webb S, Caverlee J, Pu C. Social honeypots: Making friends with a spammer near you. 2008 Presented at: Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS); August 21-22, 2008; Mountain View, CA p. 1-10 URL: https://people.engr.tamu.edu/caverlee/pubs/webb08socialhoneypots.pdf
40. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. Botornot: A system to evaluate social bots. 2016 Presented at: 25th International Conference Companion on World Wide Web; April 11-15, 2016; Montreal, Canada. [doi: 10.1145/2872518.2889302]
41. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human-bot interactions: Detection, estimation, and characterization. 2017 Presented at: International AAAI Conference on Web and Social Media; May 15-18, 2017; Montreal, Canada URL: https://arxiv.org/abs/1703.03107
42. Dickerson JP, Kagan V, Subrahmanian V. Using sentiment to detect bots on twitter: Are humans more opinionated than bots? 2014 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 17-20, 2014; Beijing, China. [doi: 10.1109/asonam.2014.6921650]
43. Stephens M, Poorthuis A. Follow thy neighbor: Connecting the social and the spatial networks on Twitter. Computers, Environment and Urban Systems 2015 Sep;53:87-95. [doi: 10.1016/j.compenvurbsys.2014.07.002]
44. Yang Z, Wilson C, Wang X, Gao T, Zhao BY, Dai Y. Uncovering social network Sybils in the wild. ACM Trans Knowl Discov Data 2014 Feb;8(1):1-29. [doi: 10.1145/2556609]
45. Pozzana I, Ferrara E. Measuring Bot and Human Behavioral Dynamics. Front Phys 2020 Apr 22;8:125. [doi: 10.3389/fphy.2020.00125]
46. Sahami M, Dumais S, Heckerman D, Horvitz E. A Bayesian approach to filtering junk email. Learning for Text Categorization: Papers from the AAAI workshop. 1998. URL: https://www.aaai.org/Papers/Workshops/1998/WS-98-05/WS98-05-009.pdf [accessed 2021-05-11]
47. Jagatic TN, Johnson NA, Jakobsson M, Menczer F. Social phishing. Commun ACM 2007 Oct;50(10):94-100 [FREE Full text] [doi: 10.1145/1290958.1290968]
48. Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks. 2010 Presented at: Annual Computer Security Applications Conference; December 6-10, 2010; Austin, TX p. 1-9. [doi: 10.1145/1920261.1920263]
49. Kudugunta S, Ferrara E. Deep neural networks for bot detection. Information Sciences 2018 Oct;467:312-322. [doi: 10.1016/j.ins.2018.08.019]
50. Mazza M, Cresci S, Avvenuti M, Quattrociocchi W, Tesconi M. RTbust: Exploiting temporal patterns for botnet detection on Twitter. 2019 Presented at: ACM Conference on Web Science; June 30, 2019; Boston, MA, USA. [doi: 10.1145/3292522.3326015]

XSL•FO
RenderX

51. Santia GC, Mujib MI, Williams JR. Detecting Social Bots on Facebook in an Information Veracity Context. 2019 Presented at: International AAAI Conference on Web and Social Media; June 11, 2019; Munich, Germany URL: https://ojs.aaai.org//index.php/ICWSM/article/view/3244
52. Yang KC, Varol O, Hui PM, Menczer F. Scalable and Generalizable Social Bot Detection through Data Selection. 2020 Presented at: AAAI Conference on Artificial Intelligence; February 7-12, 2020; New York, NY p. 1096-1103. [doi: 10.1609/aaai.v34i01.5460]
53. Treen KM, Williams HT, O'Neill SJ. Online misinformation about climate change. WIREs Clim Change 2020 Jun 18;11(5):e665 [FREE Full text] [doi: 10.1002/wcc.665]
54. Shahi GK, Dirkson A, Majchrzak TA. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021 Mar;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
55. Dwoskin E, Timberg C. Misinformation dropped dramatically the week after Twitter banned Trump and some allies. The Washington Post. 2021 Jan 16. URL: https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/ [accessed 2021-05-09]
56. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
57. Wang P, Angarita R, Renna I. Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. 2018 Presented at: Companion Proceedings of The Web Conference 2018; April 23-27, 2018; Lyon, France p. 1557-1561. [doi: 10.1145/3184558.3191610]
58. Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 2019 Jan 2;10(1):1. [doi: 10.1038/s41467-018-07761-2]
59. Ferrara E. Disinformation and social bot operations in the run up to the 2017 French presidential election. First Monday 2017 Jul 31;22(8):1. [doi: 10.5210/fm.v22i8.8005]
60. Howard PN. How Political Campaigns Weaponize Social Media Bots. IEEE Spectrum. 2018. URL: https://spectrum.ieee.org/computing/software/how-political-campaigns-weaponize-social-media-bots [accessed 2021-05-11]
61. Mueller RS. Report on the investigation into Russian interference in the 2016 presidential election. US Department of Justice. URL: https://www.justice.gov/storage/report.pdf [accessed 2021-05-09]
62. Twitter. Update on Twitter's review of the 2016 US election. Twitter Blog. 2018 Jan 19. URL: https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.html [accessed 2021-05-09]
63. Rheault L, Musulan A. Efficient detection of online communities and social bot activity during electoral campaigns. Journal of Information Technology & Politics 2021 Feb 02:1-14. [doi: 10.1080/19331681.2021.1879705]
64. Howard PN, Woolley S, Calo R. Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. Journal of Information Technology & Politics 2018 Apr 11;15(2):81-93 [FREE Full text] [doi: 10.1080/19331681.2018.1448735]
65. Assenmacher D, Clever L, Frischlich L, Quandt T, Trautmann H, Grimme C. Demystifying Social Bots: On the Intelligence of Automated Social Media Actors. Social Media + Society 2020 Sep 01;6(3):205630512093926. [doi: 10.1177/2056305120939264]
66. Varol O. Bot Repository. URL: https://botometer.osome.iu.edu/bot-repository/index.html [accessed 2020-11-20]
67. Giorgi S, Sap M. TwitterMySQL. URL: https://github.com/dlatk/TwitterMySQL [accessed 2020-11-20]
68. Al-Rawi A, Shukla V. Bots as Active News Promoters: A Digital Analysis of COVID-19 Tweets. Information 2020 Sep 27;11(10):461. [doi: 10.3390/info11100461]
69. Gilani Z, Farahbakhsh R, Tyson G, Wang L, Crowcroft J. Of bots and humans (on Twitter). 2017 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; July 31, 2017; Sydney, Australia. [doi: 10.1145/3110025.3110090]
70. Shi W, Liu D, Yang J, Zhang J, Wen S, Su J. Social Bots' Sentiment Engagement in Health Emergencies: A Topic-Based Analysis of the COVID-19 Pandemic Discussions on Twitter. Int J Environ Res Public Health 2020 Nov 23;17(22):8701. [doi: 10.3390/ijerph17228701] [Medline: 33238567]
71. Allyn B. Researchers: Nearly half of accounts tweeting about coronavirus are likely bots. NPR. 2020 May 20. URL: https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots [accessed 2021-05-09]
72. Roth Y, Pickles N. Bot or not? The facts about platform manipulation on Twitter. Twitter Blog. 2020 May 18. URL: https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.html [accessed 2021-05-09]
73. Ferrara E. What types of COVID-19 conspiracies are populated by Twitter bots? First Monday 2020 May 19;25(6):1. [doi: 10.5210/fm.v25i6.10633]
74. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
75. Allem J, Escobedo P, Dharmapuri L. Cannabis Surveillance With Twitter Data: Emerging Topics and Social Bots. Am J Public Health 2020 Mar;110(3):357-362. [doi: 10.2105/AJPH.2019.305461] [Medline: 31855475]

XSL•FO
RenderX

76. Mehta B, Salmon J, Ibrahim S. Potential Shortages of Hydroxychloroquine for Patients with Lupus During the Coronavirus Disease 2019 Pandemic. JAMA Health Forum 2020 Apr 10;1(4):e200438. [doi: 10.1001/jamahealthforum.2020.0438]
77. Lupus Research Alliance. COVID-19 Caused Hydroxychloroquine Issues for Third of Lupus Patients, New LRA Survey Finds. 2020 May 28. URL: https://www.lupusresearch.org/covid-19-caused-hydroxychloroquine-issues-for-third-of-lupus-patients-new-lra-survey-finds/ [accessed 2021-05-09]
78. Skipper CP, Pastick KA, Engen NW, Bangdiwala AS, Abassi M, Lofgren SM, et al. Hydroxychloroquine in Nonhospitalized Adults With Early COVID-19: A Randomized Trial. Ann Intern Med 2020 Oct 20;173(8):623-631 [FREE Full text] [doi: 10.7326/M20-4207] [Medline: 32673060]
79. Mitjà O, Corbacho-Monné M, Ubals M, Tebe C, Peñafiel J, Tobias A, BCN PEP-CoV-2 RESEARCH GROUP. Hydroxychloroquine for Early Treatment of Adults with Mild Covid-19: A Randomized-Controlled Trial. Clin Infect Dis 2020 Jul 16:1 [FREE Full text] [doi: 10.1093/cid/ciaa1009] [Medline: 32674126]
80. National Institutes of Health. NIH halts clinical trial of hydroxychloroquine. National Institutes of Health News Releases. 2020 Jun 20. URL: https://www.nih.gov/news-events/news-releases/nih-halts-clinical-trial-hydroxychloroquine [accessed 2021-05-09]
81. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7 [FREE Full text] [doi: 10.1017/S003329172000224X] [Medline: 32513320]
82. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar 05;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
83. Roth Y, Pickles N. Updating our approach to misleading information. Twitter Blog. 2020 May 11. URL: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html [accessed 2021-05-09]
84. Statt N. Facebook will now show a warning before you share articles about COVID-19. The Verge. 2020 Aug 12. URL: https://www.theverge.com/2020/8/12/21365305/facebook-covid-19-warning-notification-post-misinformation [accessed 2021-05-09]
85. Timberg C, Dwoskin E. Twitter is sweeping out fake accounts like never before, putting user growth at risk. The Washington Post. 2018 Jul 09. URL: https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/ [accessed 2021-05-09]
86. Facebook. Community Standards Enforcement Report. Facebook Transparency. URL: https://transparency.facebook.com/community-standards-enforcement [accessed 2021-05-09]
87. Kamal G. California's BOT Disclosure Law, SB 1001, now in effect. The National Law Review. 2019 Jul 15. URL: https://www.natlawreview.com/article/california-s-bot-disclosure-law-sb-1001-now-effect [accessed 2021-05-09]
88. Frazin R. Feinstein introduces bill to prohibit campaigns from using social media bots. The Hill. 2019 Jul 16. URL: https://thehill.com/policy/cybersecurity/453336-dem-senator-introduces-bill-to-prohibit-campaigns-from-using-bots [accessed 2021-05-09]
89. Kang C, Fandos N, Issac M. Tech executives are contrite about election meddling, but make few promises on Capitol Hill. The New York Times. 2017 Oct 31. URL: https://www.nytimes.com/2017/10/31/us/politics/facebook-twitter-google-hearings-congress.html [accessed 2021-05-09]
90. Gorwa R, Guilbeault D. Unpacking the Social Media Bot: A Typology to Guide Research and Policy. Policy & Internet 2018 Aug 10;12(2):225-248. [doi: 10.1002/poi3.184]
91. Samples J. Why the government should not regulate content moderation of social media. Cato Institute. 2019 Apr 09. URL: https://www.cato.org/publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media [accessed 2021-05-09]
92. Crews Jr CW. The Case against Social Media Content Regulation: Reaffirming Congress' Duty to Protect Online Bias, 'Harmful Content,' and Dissident Speech from the Administrative State. SSRN 2020 Jun 28;Competitive Enterprise Institute, Issue Analysis:1-34 [FREE Full text]
93. Gillespie T. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. New Haven, CT: Yale University Press; 2018.
94. Ibrahim Y. The breastfeeding controversy and Facebook: Politicization of image, privacy and protest. International Journal of E-Politics (IJEP) 2010;1(2):16-28. [doi: 10.4018/jep.2010040102]
95. Dixon L, Li J, Sorensen J, Thain N, Vasserman L. Measuring and mitigating unintended bias in text classification. 2018 Presented at: AAAI/ACM Conference on AI, Ethics, and Society; 2018; New Orleans, LA p. 67-73. [doi: 10.1145/3278721.3278729]
96. Sap M, Card D, Gabriel S, Choi Y, Smith N. The risk of racial bias in hate speech detection. 2019 Presented at: Annual Meeting of the Association for Computational Linguistics; July 2019; Florence, Italy p. 1668-1678. [doi: 10.18653/v1/p19-1163]
97. Zhao J, Xiong L, Jayashree P. Dual-agent GANs for photorealistic and identity preserving profile face synthesis. 2017 Presented at: Advances in Neural Information Processing Systems; December 4, 2017; Long Beach, CA, USA p. 66-76

XSL•FO
RenderX

URL: http://papers.neurips.cc/paper/6612-dual-agent-gans-for-photorealistic-and-identity-preserving-profile-face-synthesis.pdf

98. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P. Language models are few-shot learners. 2020 Presented at: Advances in Neural Information Processing Systems; December 6, 2020; Virtual p. 1877-1901 URL: https://arxiv.org/abs/2005.14165

Abbreviations

API: application programming interface

Edited by R Kukafka, C Basch; submitted 04.01.21; peer-reviewed by A Agarwal, D Yeung, N Martinez-Martin; comments to author 16.02.21; revised version received 04.03.21; accepted 16.04.21; published 20.05.21.

Please cite as:
Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz HA, Epstein DH, Leggio L, Curtis B
Bots and Misinformation Spread on Social Media: Implications for COVID-19
J Med Internet Res 2021;23(5):e26933
URL: https://www.jmir.org/2021/5/e26933
doi:10.2196/26933
PMID:33882014

©McKenzie Himelein-Wachowiak, Salvatore Giorgi, Amanda Devoto, Muhammad Rahman, Lyle Ungar, H Andrew Schwartz, David H Epstein, Lorenzo Leggio, Brenda Curtis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

A Real-World Rheumatology Registry and Research Consortium: The German RheumaDatenRhePort (RHADAR) Registry

Stefan Kleinert[1,2], MD; Peter Bartz-Bazzanella[3], MD; Cay von der Decken[3,4], MD; Johannes Knitza[5], MD; Torsten Witte[6], MD; Sándor P Fekete[7], Prof Dr; Matthias Konitzny[7], MSc; Alexander Zink[8], MD, MPH, PhD; Georg Gauler[9], MD; Patrick Wurth[9], MD; Peer Aries[10], MD; Kirsten Karberg[11], MD; Christoph Kuhn[12], MD; Florian Schuch[1], MD; Susanna Späthling-Mestekemper[13], MD; Wolfgang Vorbrüggen[14], MD; Matthias Englbrecht[15], Dr; Martin Welcker[16], MD; RHADAR Group[17]

[1]Praxisgemeinschaft Rheumatologie-Nephrologie, Erlangen, Germany

[2]Medizinische Klinik 3, Rheumatology/Immunology, Universitätsklinik Würzburg, Würzburg, Germany

[3]Medizinisches Versorgungszentrum Stolberg, Stolberg, Germany

[4]Klinik für Internistische Rheumatologie, Rhein-Maas-Klinikum, Würselen, Germany

[5]Department of Internal Medicine 3 – Rheumatology and Immunology, Friedrich-Alexander-University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany

[6]Department of Rheumatology and Clinical Immunology, Hanover Medical School, Hanover, Germany

[7]Department of Computer Science, TU Braunschweig, Braunschweig, Germany

[8]Department of Dermatology and Allergy, School of Medicine, Technical University of Munich, Munich, Germany

[9]Rheumatologische Schwerpunktpraxis, Osnabrück, Germany

[10]Rheumatologie im Struenseehaus, Hamburg, Germany

[11]Praxis für Rheumatologie und Innere Medizin, Berlin, Germany

[12]Praxis für Rheumatologie, Karlsruhe, Germany

[13]Rheumatologist in Group Practice, Munich, Germany

[14]Verein zur Förderung der Rheumatologie eV, Würselen, Germany

[15]Freelance Healthcare Data Scientist, Eckental, Germany

[16]Medizinisches Versorgungszentrum für Rheumatologie Dr M Welcker GmbH, Planegg, Germany

[17]see Acknowledgments

Corresponding Author:
Stefan Kleinert, MD
Praxisgemeinschaft Rheumatologie-Nephrologie
Möhrendorferstr 1c
Erlangen
Germany
Phone: 49 9131 890031
Fax: 49 9131 890050
Email: stefan.kleinert@pgrn.de

Abstract

Real-world data are crucial to continuously improve the management of patients with rheumatic and musculoskeletal diseases (RMDs). The German RheumaDatenRhePort (RHADAR) registry encompasses a network of rheumatologists and researchers in Germany providing pseudonymized real-world patient data and allowing timely and continuous improvement in the care of RMD patients. The RHADAR modules allow automated anamnesis and adaptive coordination of appointments regarding individual urgency levels. Further modules focus on the collection and integration of electronic patient-reported outcomes in between consultations. The digital RHADAR modules ultimately allow a patient-centered adaptive approach to integrated medical care starting as early as possible in the disease course. Such a closed-loop system consisting of various modules along the whole patient pathway enables comprehensive and timely patient management in an unprecedented manner.

(J Med Internet Res 2021;23(5):e28164) doi:10.2196/28164

XSL•FO
RenderX

KEYWORDS

registry; rheumatology; real-world data; symptom checker; patient-reported outcomes

Current Challenges in Rheumatology

The efficacy and safety of treatments for patients with rheumatic and musculoskeletal diseases (RMDs) have continuously improved during previous decades. However, the combination of a declining number of rheumatologists [1] on the one hand and a rising life expectancy in developed countries on the other imposes a significant challenge to rheumatology care nowadays. The growing need for a professional evaluation of common symptoms, such as joint pain [2,3], can hardly be met by the number of available rheumatologists. Besides the lack of the latter, multiple factors contribute to the omnipresent diagnostic delay in rheumatology. Unspecific rheumatic symptoms are challenging to patients [4], general practitioners [5], and experienced rheumatologists [6] alike. Even the most advanced and innovative therapies cannot reverse the health impairment that manifests due to diagnostic delay. “Time is joint” depicts the concept of a timely diagnosis as an indispensable goal in rheumatology. Focusing on this goal could be more cost-effective and generate better outcomes than the development of new RMD drugs.

The shortage of rheumatologists translates to reduced doctor-patient time, especially for follow-up appointments. On average, patients have appointments every 3 to 6 months, while rheumatologists spend 15 minutes for each consultation. In complex diseases like RMDs, the long-term treatment effectiveness cannot be deduced from a single laboratory result like HbA_{1c} in diabetes mellitus. The current monitoring approach in rheumatology relies mainly on the collection of laboratory, clinical, and anamnestic snapshots. This “black box” monitoring approach hardly represents modern personalized medicine. Apart from diagnostic delay, timely treatment adaptions are hampered by the status-quo approach.

Rheumatology encompasses various rare diseases, with limited evidence aiding clinicians in their treatment decisions. Despite the growing number of therapeutic options and treatment guidelines, patients frequently feel like guinea pigs, with prescribed anti-inflammatory treatments not showing the expected therapeutic effect. Thus, clinical evidence supported by real-world data is needed to improve personalized evidence-based treatment decisions.

Potential of a Digital Patient Pathway in Rheumatology

The potential of digitalization in rheumatology has been depicted in recent publications [7-14]. The European League Against Rheumatism (EULAR) recommends that patients experiencing morning stiffness, joint pain, or swollen joints should see a rheumatologist no later than 6 weeks after symptom onset [15]. A symptom-based, patient-centered, digital screening approach could allow the successful implementation of this guideline. This low burden approach could effectively cut diagnostic delay. The omnipresence of smartphones in the RMD population [14] (irrespective of age) and instinctive digital symptom-based search for solutions [2,3,11,14] represent an immense potential to guide and triage patient streams effectively.

Electronic patient-reported outcomes (ePROs) allow for an individual, continuous, flexible, asynchronous, and need-adapted patient follow-up. Only few rheumatologists are currently using ePROs in clinical routine, mainly due to unawareness of suitable software solutions [13]. Similarly, only few RMD patients are aware of useful rheumatology-specific medical applications [14]. However, rheumatologists and patients are motivated to use ePROs and rheumatology-specific applications in the future [9,13,14,16]. A randomized controlled trial could show that an ePROs-based telehealth follow-up strategy can achieve similar disease control as conventional follow-up [17]. This need-adapted follow-up strategy could set precious resources free, which could thus reduce the diagnostic and monitoring delay in RMD patients. A parallel collection of ePROs may facilitate the integration of ePROs into the daily routine. Importantly, wearable-based ePROs produce continuous data sets and can accurately predict disease flares in patients with rheumatoid arthritis and axial spondyloarthritis [18].

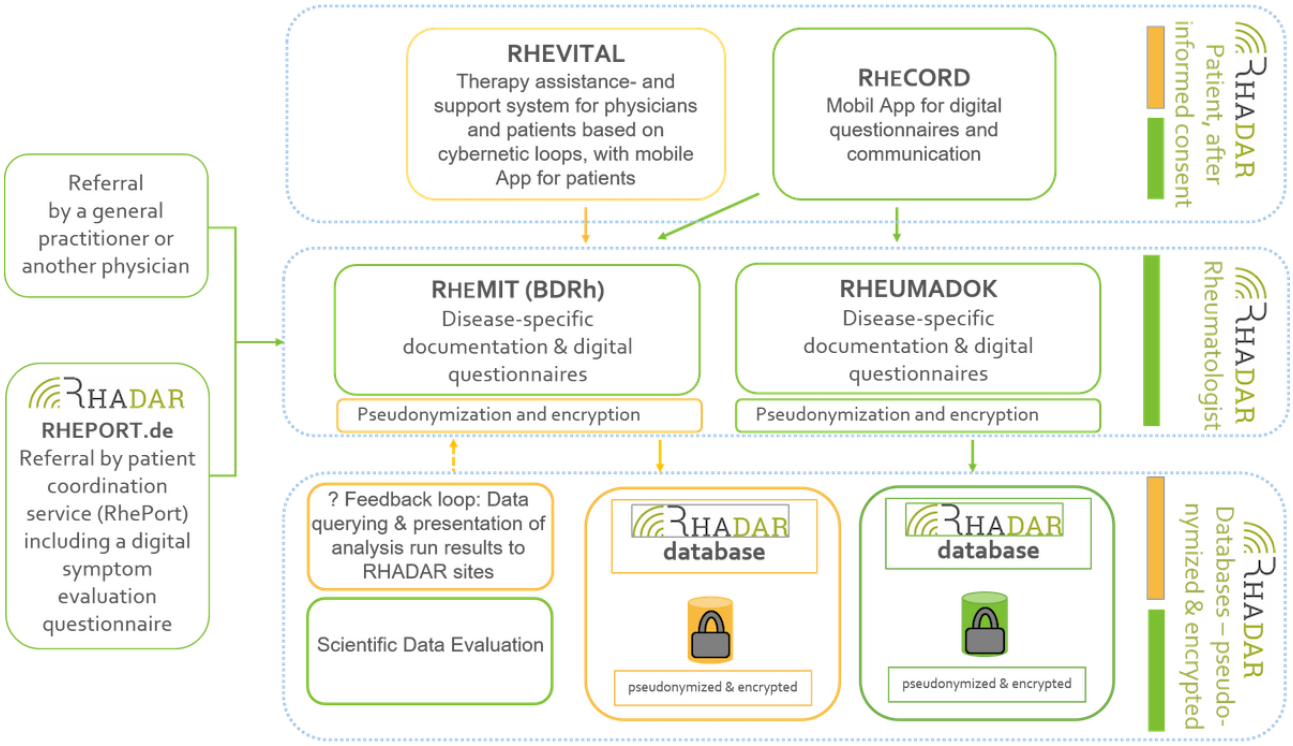
Digital registries are crucial for RMD research. Only in a collaborative effort [19] significant amounts of data can be gathered to improve knowledge among rheumatologists. Registries are urgently needed to support RMD patients’ management based on evidence. Registries have helped to provide knowledge for rare diseases, such as myositis [20], specific patient scenarios, such as pregnancy [21], and the safety of biologic therapy [22]. Unfortunately, the data collection approach of most patient registries relies on cumbersome active and redundant manual data entry, which frequently lacks precision or completeness [23]. Values can be inconsistent, out of range, or even completely made up. Most importantly, data entry and employment often stop with the end of financial support [24]. Fortunately, most RMD patients are willing to share their data for research purposes [14].

RheumaDatenRhePort Patient Pathway

The RheumaDatenRhePort (RHADAR) registry (Figure 1) is based on a framework of complementary software modules covering the patient pathway from symptom onset through the disease course.

XSL•FO
RenderX

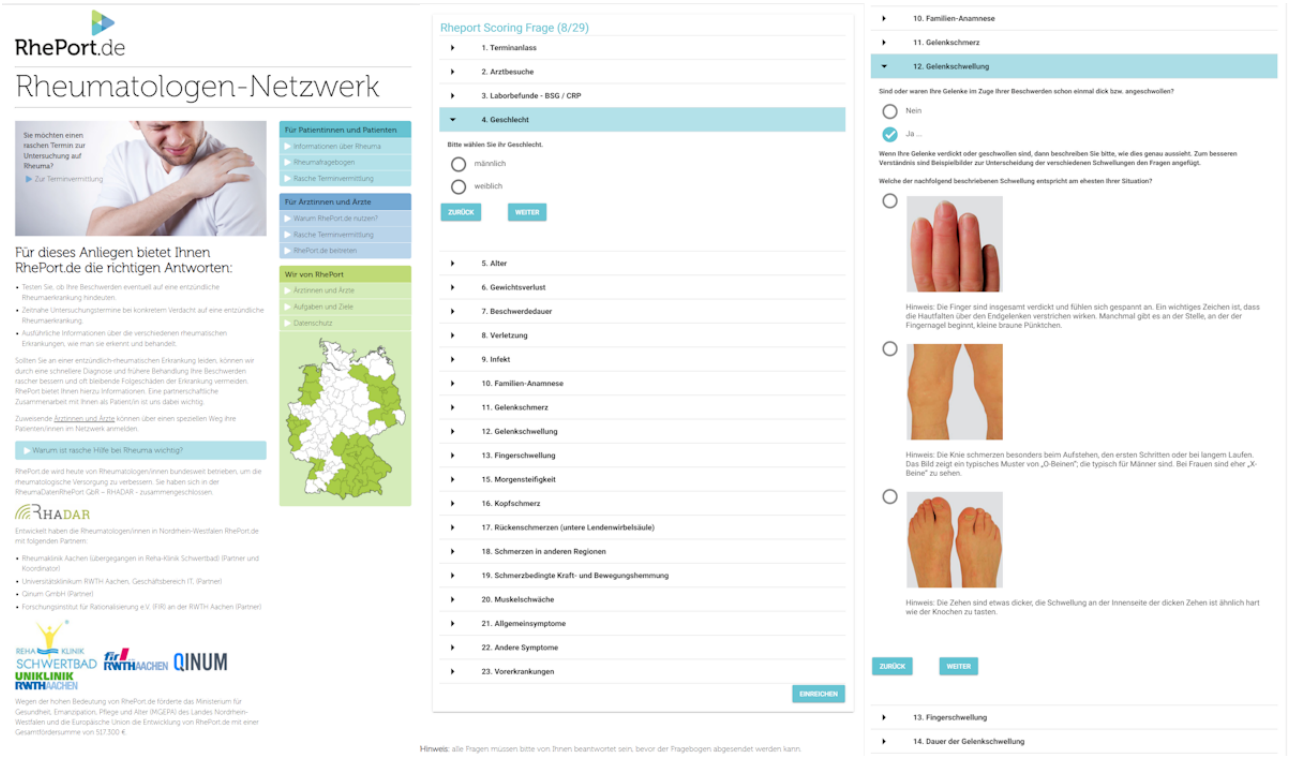
Figure 1. RHADAR patient and data pathway. Green boxes resemble established parts of the RHADAR network. Orange boxes are in preparation. RHADAR: RheumaDatenRhePort.



RhePort: Digital Gate to Rheumatology Care

RHADAR provides low burden direct digital access to a growing network of participating rheumatology sites. The gate module toward RHADAR rheumatology care is RhePort (Figure 2) [25]. Online registration is necessary to use RhePort, and users need to actively consent to the addition of their data to the RHADAR registry.

Figure 2. Screenshots of RhePort.



RhePort currently consists of a questionnaire that takes on average 9 minutes to complete [26]. It includes necessary demographic information, current symptoms, and already collected laboratory reports. Based on the patient's answers, RhePort calculates the RMD risk. The implemented algorithm refers to a weighted sum score. Individual weights were derived

XSL•FO
RenderX

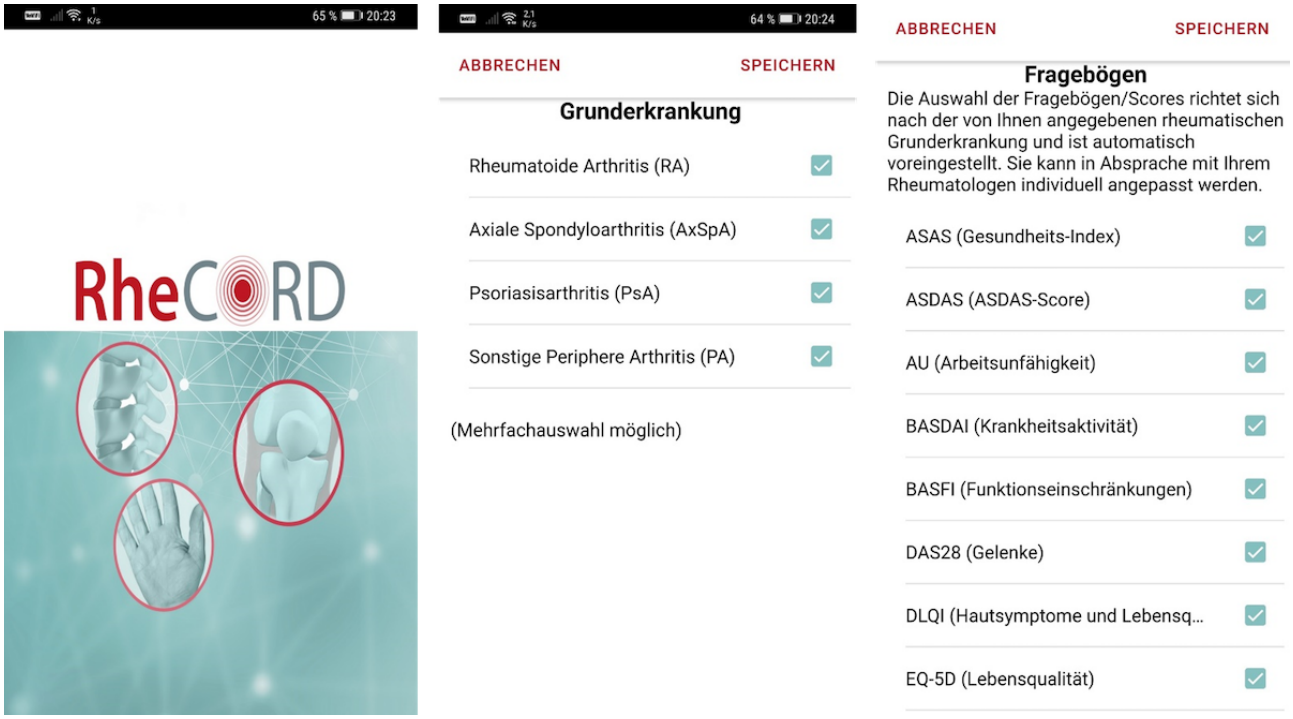
by consensus and the work experience of the RhePort founders. Based on the RhePort RMD risk score, an individual triage level is determined. Currently, there are four triage levels as follows: (1) very urgent, (2) urgent, (3) intermediate, and (4) unlikely. Based on the corresponding level, the patient gets access to locally available RHADAR rheumatology site appointments.

The questionnaire threshold focuses on sensitivity, that is, omitting false-negative ratings. Furthermore, participating rheumatology sites are encouraged to use RhePort only on a complimentary basis. Proft et al recently showed that an online self-referral strategy can successfully be used in addition to the traditional physician-based referral [27]. By including patients based on symptoms, at-risk RMD patients are included in the registry as well. In an interim analysis of a head-to-head randomized controlled trial, the diagnostic accuracy of RhePort was similar to an artificial intelligence–based symptom checker regarding the presence of RMDs [26].

RheCORD: Digital Companion App

RheCORD (Figure 3) [28] is a rheumatological medical app currently being developed to enable the collection of ePROs. RheCORD gives insights into the patient's symptom state between consultations or before an upcoming appointment on the same day. The ePROs to be collected can individually be determined by the patient [14,29] and local rheumatology site [13] via shared decision-making. The app provides multiple optional secondary features, such as medication and appointment reminders. Feature selection and app development are carried out with respect to guidelines [30,31], app analysis [31], and patient insights [14,29]. RheCORD is based on the previous CE-certified medical apps RheumaLive, AxSpALive, and PsALive. RHADAR sites provide a free soft token and password to the patient via a QR code to enable encrypted data transfer to the site.

Figure 3. Screenshots of RheCORD.

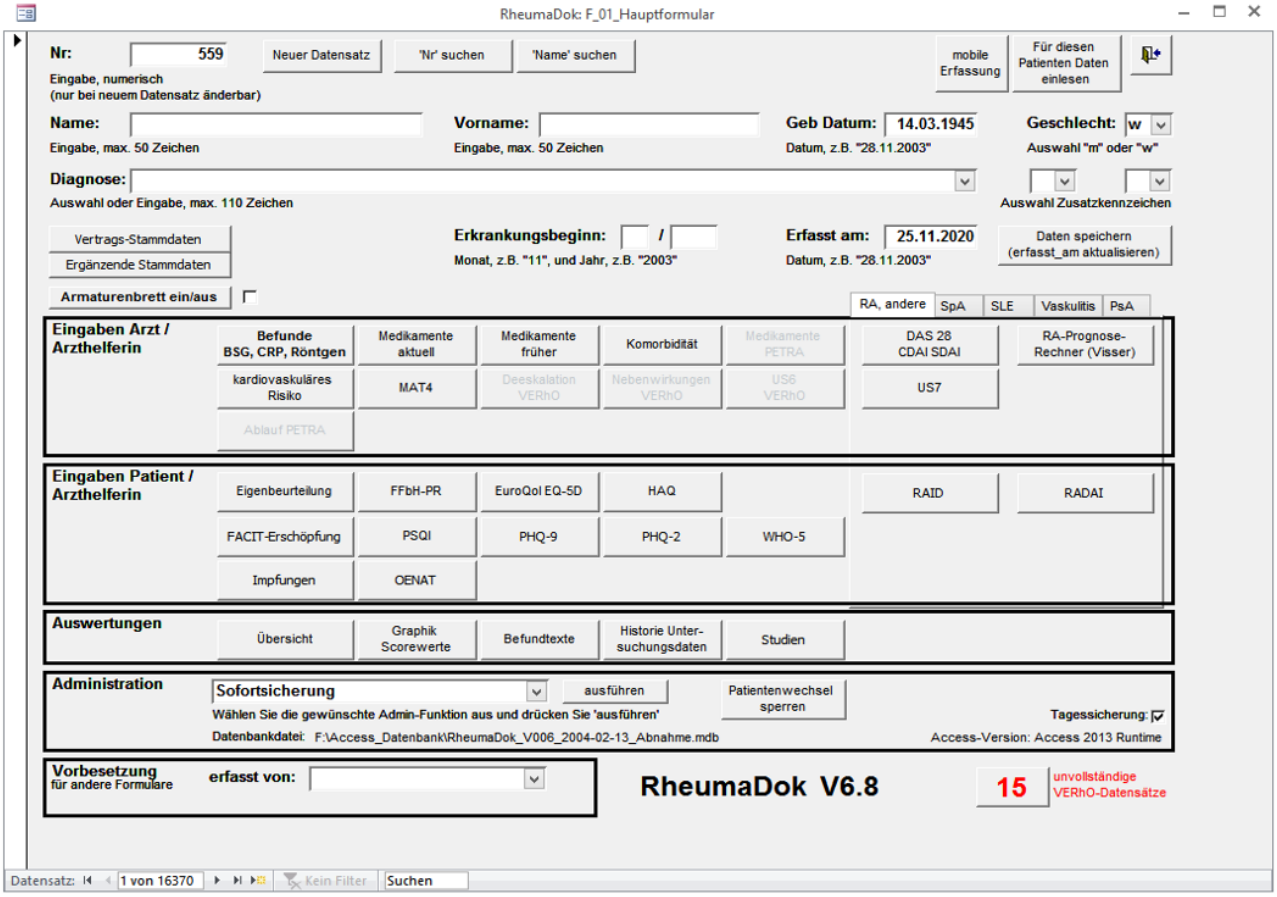


RheumaDok

RheumaDok (Figure 4) is the current core application used by the RHADAR system for patient data collection. It is a Microsoft Access (Microsoft Corp) database developed in Visual Basic for Applications and is available free of charge for rheumatologists in Germany who are members of the German Rheumatologists' Professional Association (Berufsverband Deutscher Rheumatologen [BDRh]). This database was developed by the Nils Körber und Joachim Elgas GbR. In the months ahead, RheumaDok will be replaced by RheMIT.

XSL•FO
RenderX

Figure 4. Screenshot of RheumaDok.

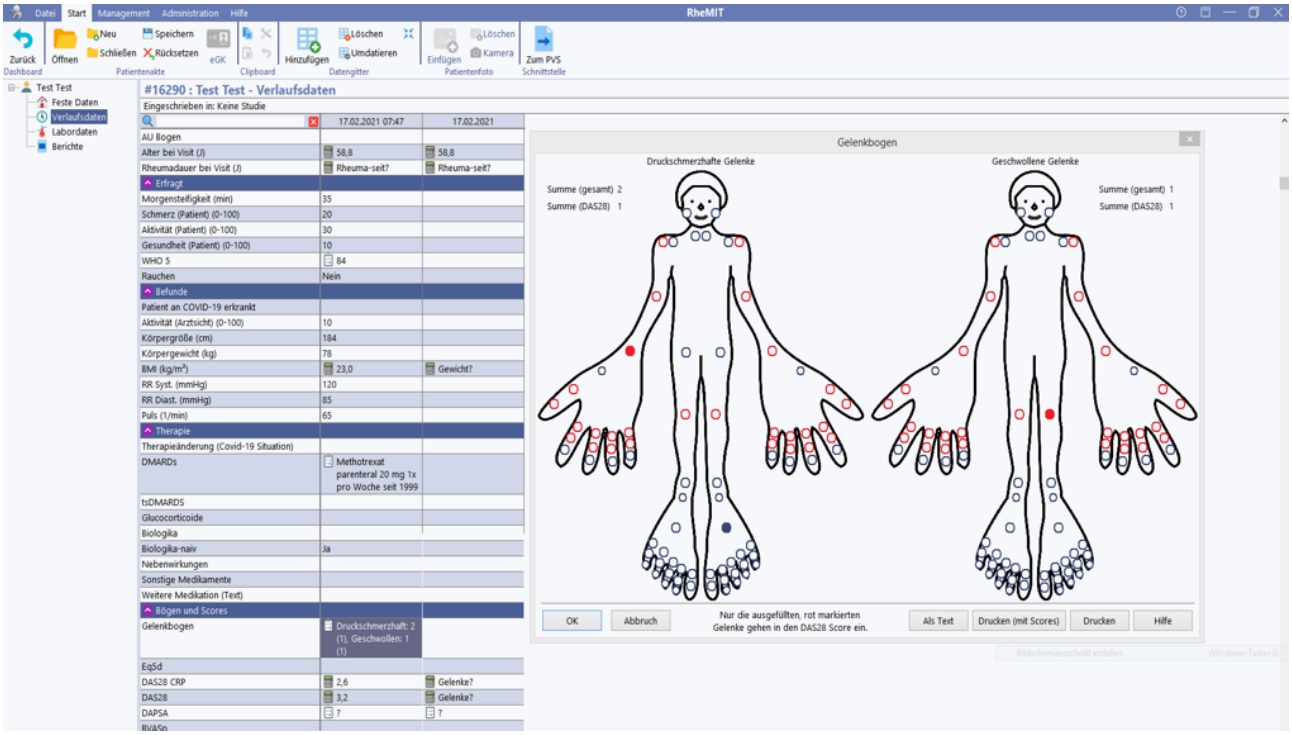


RheMIT: Clinical Documentation Software

RheMIT (Figure 5) is currently becoming the standard clinical documentation software for rheumatologists in Germany. The software has been developed and adapted by itc-ms.de for the German Rheumatologists' Professional Association (BDRh) [7]. The software offers standardized entry masks to document the patient's disease course at each consultation, including standard disease activity measures, general or disease-specific PROs, demographic characteristics, comorbidities, and current as well as previous antirheumatic treatments. If patients provide their consent, pseudonymized information will be uploaded to the RHADAR registry.

XSL•FO
RenderX

Figure 5. Screenshot of RheMIT.



RHADAR Registry

All information from RheCORD and local RheumaDok database instances are pseudonymized at each participating site. Subsequently, data are merged in a mutual database. The responsible data protection authority approved the current RHADAR database environment to ensure data privacy according to legislation. Patients participating in the RHADAR network have to complete an informed consent form before their initial data entry. Participation is voluntary for each patient. Participants and patients refusing to participate are treated equally, following national guidelines and rheumatologic care standards in Germany. As of September 2020, 22 rheumatologists at eight sites provided data to the RHADAR registry, including a total of 15,908 patients.

RHADAR Outlook

RHADAR software modules are continuously being improved based on new technological standards, user preferences, and scientific literature. For instance, a machine-learning approach to improve the triage performance of the RhePort algorithm is being evaluated based on current data. The implementation of an objective, standardized, and evidence-based triage algorithm seems crucial. Additionally, a complementary software module, RheVITAL [32], is under development in cooperation with the largest German patient organization, the German Rheumatism League (“Deutsche Rheuma-Liga”), to facilitate RMD disease management. Several registry analyses are being carried out and are set to directly impact future RMD management. A new database is currently under construction, and it will include data from all current and future software modules. Growing awareness of the integrative RHADAR system and its distribution will foster networking of rheumatology sites.

Acknowledgments

The authors thank all patients, partners, and collaborators for their support of the RHADAR registry. The authors would especially like to thank Monika Ronneberger, Praxedis Rapp, Jörg Wendler, Sonja Baltrusch, Florian Popp, Patrizia Sternad, Alexandra Edhofer, Ernestina Guillen Poveda, Yella C Rottländer, Sybille Kramer, Benjamin Schulte, Ramona Merkle, and Dorthe Gruber for their support in the data acquisition process and during data completion. The authors would like to thank Nils Körber, the developer of the initial RHADAR database, for his support and consultation during the project roll out; Michael Schumann, the developer of the successor RHADAR database, for his support and cooperation; and Achim Fricker, the cooperation partner developing RhePort with the founders, for his support. This study was funded by the RheumaDatenRhePort GbR, Bahnhofstr. 32, 82152 Planegg, Germany.

The members of the RHADAR Group, Planegg, Germany includes Monika Ronneberger, Praxedis Rapp, Jörg Wendler, Sonja Baltrusch, Florian Popp, Patrizia Sternad, Alexandra Edhofer, Ernestina Guillen Poveda, Yella C Rottländer, Sybille Kramer, Benjamin Schulte, Ramona Merkle, Dorthe Gruber, Nils Körber, and Achim Fricker.

XSL•FO
RenderX

Authors' Contributions

All authors wrote, reviewed, and approved the final manuscript.

Conflicts of Interest

WV, CD, SK, PB-B, and MW are members of RheumaDatenRhePort. WV, CD, and PB-B were involved in the development of Rheport. JK, AZ, TW, and SPF are members of the scientific board of RheumaDatenRhePort.

RhePort was developed and provided by the Verein zur Förderung der Rheumatologie eV, Würselen, Germany, and QINUM GmbH, Cologne, Germany. The development was supported by the government of Northrhine-Westfalia, Germany, and by the European Union (EFRE; AZ: 005-GW02-075A). The maintenance of this service has been supported by AbbVie, Novartis, Chugai, UCB, and Sanofi-Aventis. RheVITAL was developed and provided by the Verein zur Förderung der Rheumatologie eV, Würselen, Germany; QINUM GmbH, Cologne, Germany; and the Martin-Luther- Universität Halle, Halle, Germany. The development was supported by the government of Northrhine-Westfalia, Germany, and by the European Union (EFRE; Förderkennzeichen: EFRE-0800750). RheumaDok and the RHADAR database were developed and technically implemented by Nils Körber and Joachim Elgas GbR, Erlangen, Germany. RheMIT has been developed for the German Rheumatologists' Professional Association (Berufsverband Deutscher Rheumatologen) by itc-ms.de, Marburg, Germany. The app RheCORD is the property of the RheCordHealthCare GmbH. It was preceded by AxSpALive, PsALive, and RheumaLive. The latter three were sponsored by UCB Pharma GmbH, Monheim, Germany. All apps were developed and technically implemented by Star Healthcare Management GmbH, Cologne, Germany. JK has received research support from Novartis Pharma GmbH to evaluate RhePort.

GG receives speaker and consulting fees from Abbvie, Celgene, Chugai, Gilead, Novartis, and UCB. CK, KK, and SS-M are members of RheumaDatenRhePort GbR that had part of the software developed as described in this publication. FS is part of the advisory board at Abbvie, Gilead, Novartis, and Sanofi Aventis. ME has received remunerations from RheumaDatenRhePort GbR for statistical data analyses & consultation.

The RHADAR GbR has received honoraria from UCB Pharma GmbH, Sandoz Deutschland/Hexal AG, and Lilly GmbH, and research support from Novartis Pharma GmbH.

References

1. Krusche M, Sewerin P, Kleyer A, Mucke J, Vossen D, Morf H, Rheumadocs und Arbeitskreis Junge Rheumatologie (AGJR). Facharztweiterbildung quo vadis? Specialist training quo vadis? Z Rheumatol 2019 Oct;78(8):692-697. [doi: 10.1007/s00393-019-00690-5] [Medline: 31468164]
2. Dey M, Zhao SS. Google search data as a novel adjunct to patient and public involvement in rheumatology research. Rheumatol Int 2021 Apr 19;41(4):771-779 [FREE Full text] [doi: 10.1007/s00296-020-04723-0] [Medline: 33074346]
3. Kardeş S, Kuzu AS, Raiker R, Pakhchanian H, Karagülle M. Public interest in rheumatic diseases and rheumatologist in the United States during the COVID-19 pandemic: evidence from Google Trends. Rheumatol Int 2021 Feb;41(2):329-334 [FREE Full text] [doi: 10.1007/s00296-020-04728-9] [Medline: 33070255]
4. Raciborski F, Kłak A, Kwiatkowska B, Batko B, Sochocka-Bykowska M, Zoń-Giebel A, working team. Diagnostic delays in rheumatic diseases with associated arthritis. Reumatologia 2017;55(4):169-176 [FREE Full text] [doi: 10.5114/reum.2017.69777] [Medline: 29056771]
5. van der Linden MPM, le Cessie S, Raza K, van der Woude D, Knevel R, Huizinga TWJ, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis Rheum 2010 Dec 30;62(12):3537-3546 [FREE Full text] [doi: 10.1002/art.27692] [Medline: 20722031]
6. Ehrenstein B, Pongratz G, Fleck M, Hartung W. The ability of rheumatologists blinded to prior workup to diagnose rheumatoid arthritis only by clinical assessment: a cross-sectional study. Rheumatology (Oxford) 2018 Sep 01;57(9):1592-1601. [doi: 10.1093/rheumatology/key127] [Medline: 29850885]
7. Fiehn C, Baraliakos X, Edelmann E, Froschauer S, Feist E, Karberg K, et al. [Current state, goals and quality standards of outpatient care in rheumatology: position paper of the Professional Association of German Rheumatologists (BDRh)]. Z Rheumatol 2020 Oct;79(8):770-779. [doi: 10.1007/s00393-020-00872-6] [Medline: 32926218]
8. Knitza J, Callhoff J, Chehab G, Hueber A, Kiltz U, Kleyer A, et al. [Position paper of the commission on digital rheumatology of the German Society of Rheumatology: tasks, targets and perspectives for a modern rheumatology]. Z Rheumatol 2020 Aug;79(6):562-569. [doi: 10.1007/s00393-020-00834-y] [Medline: 32651681]
9. Kernder A, Morf H, Klemm P, Vossen D, Haase I, Mucke J, et al. Digital rheumatology in the era of COVID-19: results of a national patient and physician survey. RMD Open 2021 Feb;7(1) [FREE Full text] [doi: 10.1136/rmdopen-2020-001548] [Medline: 33622673]
10. Knitza J, Knevel R, Raza K, Bruce T, Eimer E, Gehring I, JPAST Group. Toward Earlier Diagnosis Using Combined eHealth Tools in Rheumatology: The Joint Pain Assessment Scoring Tool (JPAST) Project. JMIR Mhealth Uhealth 2020 May 15;8(5):e17507 [FREE Full text] [doi: 10.2196/17507] [Medline: 32348258]

XSL•FO
RenderX

11. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Aug 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
12. Hirsch MC, Ronicke S, Krusche M, Wagner AD. Rare diseases 2030: how augmented AI will support diagnosis and treatment of rare diseases in the future. Ann Rheum Dis 2020 Jun;79(6):740-743 [FREE Full text] [doi: 10.1136/annrheumdis-2020-217125] [Medline: 32209541]
13. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, Usage, and Barriers of Electronic Patient-Reported Outcomes Among German Rheumatologists: Survey Study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e18117 [FREE Full text] [doi: 10.2196/18117] [Medline: 32390592]
14. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile Health Usage, Preferences, Barriers, and eHealth Literacy in Rheumatology: Patient Survey Study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
15. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Álvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017 Jun;76(6):948-959. [doi: 10.1136/annrheumdis-2016-210602] [Medline: 27979873]
16. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. [Use of medical apps and online platforms among German rheumatologists : Results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs]. Zeitschrift für Rheumatologie 2019;78:839-846. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
17. de Thurah A, Stengaard-Pedersen K, Axelsen M, Fredberg U, Schougaard LMV, Hjollund NHI, et al. Tele-Health Followup Strategy for Tight Control of Disease Activity in Rheumatoid Arthritis: Results of a Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018 Mar 23;70(3):353-360 [FREE Full text] [doi: 10.1002/acr.23280] [Medline: 28511288]
18. Gossec L, Guyard F, Leroy D, Lafargue T, Seiler M, Jacquemin C, et al. Detection of Flares by Decrease in Physical Activity, Collected Using Wearable Activity Trackers in Rheumatoid Arthritis or Axial Spondyloarthritis: An Application of Machine Learning Analyses in Rheumatology. Arthritis Care Res (Hoboken) 2019 Oct;71(10):1336-1343. [doi: 10.1002/acr.23768] [Medline: 30242992]
19. Krusche M, Burmester GR, Knitza J. Digital crowdsourcing: unleashing its power in rheumatology. Ann Rheum Dis 2020 Sep;79(9):1139-1140 [FREE Full text] [doi: 10.1136/annrheumdis-2020-217697] [Medline: 32527863]
20. Lilleker JB, Vencovsky J, Wang G, Wedderburn LR, Diederichsen LP, Schmidt J, et al. The EuroMyositis registry: an international collaborative tool to facilitate myositis research. Ann Rheum Dis 2018 Jan;77(1):30-39 [FREE Full text] [doi: 10.1136/annrheumdis-2017-211868] [Medline: 28855174]
21. Müller-Ladner U, Fischer-Betz R. German Pregnancy Register "Rhekiss". Dtsch Arztebl Int 2018 Feb 16;115(7):112 [FREE Full text] [doi: 10.3238/arztebl.2018.0112a] [Medline: 29510821]
22. Strangfeld A, Zink A. Sicherheit unter Biologika – Ergebnisse aus dem deutschen Biologika-Register RABBIT [Safety of biologic therapy - results from the German biologics register RABBIT]. Dtsch Med Wochenschr 2014 Sep;139(37):1817-1820. [doi: 10.1055/s-0034-1370252] [Medline: 25180996]
23. Managing Missing Data in Patient Registries. Agency for Healthcare Research and Quality. 2018. URL: https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/white-paper-2016-2 [accessed 2021-05-17]
24. Garrison HH, Deschamps AM. NIH research funding and early career physician scientists: continuing challenges in the 21st century. FASEB J 2014 Mar;28(3):1049-1058 [FREE Full text] [doi: 10.1096/fj.13-241687] [Medline: 24297696]
25. RhePort. URL: www.rheport.de [accessed 2021-05-13]
26. Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Ab1346-Hpr Real-World Effectiveness and Perceived Usefulness of Symptom Checkers in Rheumatology: Interim Report from the Prospective Multicenter Better Study. Ann Rheum Dis 2020 Jun 02;79(Suppl 1):1961. [doi: 10.1136/annrheumdis-2020-eular.1604]
27. Proft F, Spiller L, Redeker I, Protopopov M, Rodriguez VR, Muche B, et al. Comparison of an online self-referral tool with a physician-based referral strategy for early recognition of patients with a high probability of axial spa. Semin Arthritis Rheum 2020 Oct;50(5):1015-1021. [doi: 10.1016/j.semarthrit.2020.07.018] [Medline: 32911279]
28. RheCORD. URL: www.rhecord.de [accessed 2021-05-13]
29. Nowell WB, Gavigan K, Kannowski CL, Cai Z, Hunter T, Venkatachalam S, et al. Which patient-reported outcomes do rheumatology patients find important to track digitally? A real-world longitudinal study in ArthritisPower. Arthritis Res Ther 2021 Feb 10;23(1):53 [FREE Full text] [doi: 10.1186/s13075-021-02430-0] [Medline: 33568191]
30. Najm A, Nikiphorou E, Kostine M, Richez C, Pauling JD, Finckh A, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self-management in people living with rheumatic and musculoskeletal diseases. RMD Open 2019;5(2):e001014 [FREE Full text] [doi: 10.1136/rmdopen-2019-001014] [Medline: 31565245]
31. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
32. RheVITAL. URL: www.rhevital.de [accessed 2021-05-13]

XSL•FO
RenderX

Abbreviations

BDRh: Berufsverband Deutscher Rheumatologen (German Rheumatologists' Professional Association)
ePROs: electronic patient-reported outcomes
RHADAR: RheumaDatenRhePort
RMD: rheumatic and musculoskeletal disease

Edited by R Kukafka; submitted 23.02.21; peer-reviewed by A Rovetta; comments to author 17.03.21; revised version received 19.03.21; accepted 30.04.21; published 20.05.21.

Please cite as:
Kleinert S, Bartz-Bazzanella P, von der Decken C, Knitza J, Witte T, Fekete SP, Konitzny M, Zink A, Gauler G, Wurth P, Aries P, Karberg K, Kuhn C, Schuch F, Späthling-Mestekemper S, Vorbrüggen W, Englbrecht M, Welcker M, RHADAR Group
A Real-World Rheumatology Registry and Research Consortium: The German RheumaDatenRhePort (RHADAR) Registry
J Med Internet Res 2021;23(5):e28164
URL: https://www.jmir.org/2021/5/e28164
doi:10.2196/28164
PMID:34014170

©Stefan Kleinert, Peter Bartz-Bazzanella, Cay von der Decken, Johannes Knitza, Torsten Witte, Sándor P Fekete, Matthias Konitzny, Alexander Zink, Georg Gauler, Patrick Wurth, Peer Aries, Kirsten Karberg, Christoph Kuhn, Florian Schuch, Susanna Späthling-Mestekemper, Wolfgang Vorbrüggen, Matthias Englbrecht, Martin Welcker, RHADAR Group. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Artificial Intelligence Can Improve Patient Management at the Time of a Pandemic: The Role of Voice Technology

Tomasz Jadczyk[1,2], MD, PhD; Wojciech Wojakowski[1], MD, PhD; Michal Tendera[1], MD; Timothy D Henry[3], MD; Gregory Egnaczyk[3], MD; Satya Shreenivas[3], MD

[1]Department of Cardiology and Structural Heart Diseases, Medical University of Silesia, Katowice, Poland

[2]Interventional Cardiac Electrophysiology Group, International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic

[3]The Carl and Edyth Lindner Center for Research and Education, The Christ Hospital, Cincinnati, OH, United States

Corresponding Author:
Tomasz Jadczyk, MD, PhD
Department of Cardiology and Structural Heart Diseases
Medical University of Silesia
Ziolowa 45-47
Katowice, 40-635
Poland
Phone: 48 512 099 211
Email: tomasz.jadczyk@gmail.com

Abstract

Artificial intelligence–driven voice technology deployed on mobile phones and smart speakers has the potential to improve patient management and organizational workflow. Voice chatbots have been already implemented in health care–leveraging innovative telehealth solutions during the COVID-19 pandemic. They allow for automatic acute care triaging and chronic disease management, including remote monitoring, preventive care, patient intake, and referral assistance. This paper focuses on the current clinical needs and applications of artificial intelligence–driven voice chatbots to drive operational effectiveness and improve patient experience and outcomes.

(J Med Internet Res 2021;23(5):e22959) doi:10.2196/22959

KEYWORDS

artificial intelligence; conversational agent; COVID-19; virtual care; voice assistant; voice chatbot

Review of Voice Assistants in Health Care

A cutting-edge development in the field of artificial intelligence (AI) and machine learning has enabled verbal communication interfaces between users and voice assistants (VAs), which are synonymously termed “voice chatbots” or “conversational agents.” VAs, exemplified by Apple’s Siri, Amazon Alexa, and Google Assistant, are a software layer embedded in a standalone smart speaker, such as Home Pod, Amazon Echo, Google Home, or smartphones, allowing for the interpretation of human speech. Technically, conversational agents are cloud-based services performing speech-to-text and text-to-speech tasks that are initiated by a person evoking a wake-word followed by a voice command [1]. The implementation of VAs in health care has the potential to support the delivery of care in a routine clinical setting, especially when functionalized by a computerized clinical decision support system (CDSS). These robotic process automation (RPA) chatbots have the decision-making ability of a human health care professional to perform rule-based tasks (ie, digital patient triaging) [2]. Voice technology has already been tested in various applications to support everyday clinical activities [3-12]. Its applications include the following:

1. Education-level services answering frequently asked questions based on a knowledge database (eg, first aid instructions).
2. Optimization of processes (eg, medication reminders, prescription refills, appointment scheduling, bedside assistants, and paperless documentation).
3. Patient support with personalized rule-based clinical instructions (eg, instructions to reduce carbohydrate intake among patients with diabetes mellitus).
4. Data collection services (eg, collection of patient-reported outcomes, biometric tracking, and identification of health status changes exemplified by the collection of medical history or remote home monitoring; classified by the Food and Drug Administration [FDA] as Medical Device Data Systems [13,14]).

XSL•FO
RenderX

5. Medical device–grade solutions designed to diagnose, treat, cure, mitigate, or prevent disease, termed "Software as a Medical Device" by the FDA [15], combining voice interface and CDSS.

Education-level VA services have been implemented by health care organizations and provide guidelines, instructions, and navigation for patients, such as WebMD [16], Cleveland Clinic's Tip of the Day [17], Boston Children's KidsMD [18], Mayo Clinic's First Aid [19], American Red Cross' First Aid [20], Mayo Clinic's News Network [21], Boston Children's My Children's Enhanced Recovery After Surgery [22], Ohio Health [23], or New Hanover Regional Medical Center [24]. For example, Mayo Clinic's First Aid application deployed on Amazon Alexa provides self-care instructions for everyday mishaps. Through a voice interface, users can ask for guidance on how to treat fever or what to do in case of spider bites or burns. These examples confirm the applicability of conversational systems for responding to health-related search activities. Moreover, the Livongo Blood Sugar Lookup app is designed for patients with diabetes mellitus, which supports them to keep track of their blood sugar levels logged via Livongo meter [25]. This solution provides recent blood glucose readings through verbal input.

The optimization of processes and workflow with VAs has proved useful for patients in the pharmacotherapy management at home. The Alexa-based Giant Eagle Pharmacy app allows users to set reminders and helps refill prescriptions through home delivery services [26]. Similar functionality is available through the Express Scripts app [27]. Furthermore, Swedish Health Connect by Providence St Joseph Health allows for scheduling of medical visits by suggesting the next available appointments near the patients' homes [28]. Similarly, the Atrium Health app provides information about the nearest urgent care service and hospital wait times, hours of operation, and contact details, thus allowing patients to schedule a same-day visit with a care provider [29]. OrbitaASSIST voice technology was also implemented as a bedside assistant to optimize communication with care teams and smart routing of requests [30]. Moreover, researchers from the Nationwide's Children Hospital (Columbus, Ohio) designed the SpeakHealth voice interactive service for the care coordination of children with complex medical problems [3], which confirms the applicability of voice-enabled technology in pediatrics. Conversational agents have been also shown to optimize hospital operations at Cedars-Sinai Medical Center (Los Angeles, California), supporting health care professionals in time-consuming paperwork tasks and automating medical data collection and documentation through the electronic health record (EHR)–integrated CardioCube voice app (accuracy=97.5%) [7].

Personalized clinical instructions are an important component of a holistic medical approach allowing for the continuity and coordination of care. Answers by Cigna tracks patient incentive programs, provides wellness tips, and enables health coach programs to navigate treatment plans [31]. OrbitaCONNECT provides a virtual health assistant for chronic, pre-, and postvisit care continuity [32]. Furthermore, with Talkspace Alexa skill, users can access mental health assessments and a library of mental health tools [33].

End-to-end solutions powered by voice AI can be used for routine clinical care. CardioCube allows for the collection of patient-reported outcomes and biometric data captured at patients' homes. At the appointed time, the voice assistant has a medical conversation with the patient by asking him/her, "What's your blood pressure?" Accordingly, the patient measures his/her blood pressure by using a standard monitor and reads out the results to the voice assistant. CardioCube also asks about dyspnea, quality of life, or prompts for tasks including, "Let's check your ischemic and bleeding risk again," in reference to CHA_2DS_2-VASc/HAS-BLED scores. Afterward, CardioCube automatically transmits the results to a proprietary server integrated with the EHR system and red-flags any alarming reports. This solution, which complies with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the General Data Protection Regulation, was validated at Cedars Sinai Medical Center [7] and classified by the FDA as a Medical Device Data System. CardioCube was also implemented for remote home monitoring of adult patients with heart failure and diabetes at the Family Care Network (Bellingham, Washington) (Multimedia Appendix 1 and Figure 1 exemplify the utility of the CardioCube voice app for patients with diabetes mellitus and a medical report automatically generated from the conversation with the virtual assistant, respectively).

XSL•FO
RenderX

Figure 1. Medical report generated automatically from the artificial intelligence–driven CardioCube voice app for patients with diabetes.

#	Question	Patient response	Date
1.	In the past week, have you missed any doses of your medication?	NO	03/04/20 8:56 AM
2.	Are you needing a medication refill?	NO	03/04/20 8:56 AM
3.	Do you have any medication related questions that you need your care team to answer?	YES	03/04/20 8:56 AM
4.	A caring reminder, eating more carbohydrates increases your blood sugar. All sugary foods contain carbohydrates, as do bread, rice, pasta and potatoes. Have you been carefully managing your carbohydrate intake in the past week?	YES	03/04/20 8:56 AM
5.	And what about exercise, how many times in the past week have you exercised?	2	03/04/20 8:57 AM
6.	As for this past week, were you able to check your sugar levels with a glucometer?	YES	03/04/20 8:57 AM
7.	And how many times in the past week did you check your blood sugar level?	7	03/04/20 8:57 AM
8.	Were the majority of your readings in a good range?	YES	03/04/20 8:57 AM

Kind regards,
CardioCube Team

support@cardiocube.com
cardiocube.com
1411 4th Ave, Seattle, WA 98101

Furthermore, focusing on cardiovascular diseases, Shara et al [34] at Medstar Health Research Institute conducted a clinical trial using Amazon Alexa as an automated personal health care assistant for patients with heart failure to optimize clinical care (trial registration# NCT03707275). Outcome measures included the change in the number of hospitalizations and medication adherence during a 3-month follow-up period (upon study completion, when awaiting for publication of the results). VAs have been also implemented in other fields of medicine. For example, Beaman et al [5] at the Oklahoma State University Center for Health Sciences initiated a study to examine whether verbal responses collected through Amazon Alexa is effective in capturing participant depression levels, using Patient Health Questionnaire-9 (trial registration# NCT04609267) [5].

The most advanced medical software solutions are defined to be SaMD, which, under the Federal Food, Drug, and Cosmetic Act include services designed to diagnose, treat, cure, mitigate, or prevent disease. Interestingly, regulatory authorities tend to facilitate the implementation of chatbot technology into clinical practice, especially in the context of the current pandemic. For example, the Danish COVID-19 RPA triaging chatbot that incorporates if-then branching logic was evaluated as being of “lower risk” by the FDA despite featuring a diagnostic component when the Federal Food, Drug, and Cosmetic Act was not enforced [35]. To our knowledge, none of the current VA-deployed health care apps are classified as SaMD.

Gaps in Health Care Delivery Exposed by the COVID-19 Pandemic

The surge in COVID-19 cases has placed unprecedented strain on health care systems, requiring adjustments in treatment delivery to patients. Despite the fact that the traditional clinical approach was partially substituted with web-based visits, the mismatch between demand and resources is a realistic challenge. The capacity of health care systems to adjust is limited by the incremental rate at which systems can grow by training new health care providers and reorganization of the structure [36]. However, the exponential increase in COVID-19 cases has highlighted gaps in the current organizational systems at a global scale [37]. As a result of the number of COVID-19 cases, health care providers have been required to work on the front line, shifting the human resources available for routine care services. Clinical institutions underwent extraordinary reorganizations to accommodate for the surge in COVID-19 cases. The focus was on emergency care, scaling up beds in intensive care units, and reassigning roles among the clinical staff. Parallelly, nonurgent procedures and elective surgeries were postponed. Furthermore, shortages of personal protective equipment, especially in the early phase of the pandemic, exposed health care workers to the infection and the incapability to combat the COVID-19 pandemic [38-40]. Subsequently, we observed an unprecedented health care crisis having a direct and indirect impact on medical patients with and those without COVID-19. Combined with a high number of deaths and serious illness among patients with COVID-19 [41], there were reductions in accident, emergency, and hospital admissions for urgent conditions, such as myocardial infarctions [42], which reflects patients who decided not to seek medical care owing to the fear of becoming infected. Disruptions due to the pandemic affected people with chronic conditions who could not access routine medical services. It let to postponed elective procedures, on-site visits, and reduced rates of hospitalization during the COVID-19 emergency. This will likely result in an extensive workload in the postpandemic period.

XSL•FO
RenderX

The health care crisis magnified the problem among socioeconomic statuses and racial groups [43]. Furthermore, as pointed out by the World Health Organization, public health gaps impacted the security and economic situation [44], thus revealing deep underlying problems in the insurance coverage system in the United States. A sudden wave of unemployment caused many people to lose employer-sponsored insurance coverage, thus limiting access to care in low-income populations. The COVID-19 pandemic also impacted employers who faced several challenges when operating in a difficult organizational situation. Some institutions including hospitals were required to screen all employees and visitors for COVID-19 symptoms prior to entrance. For example, University of California San Francisco Health was posed with a tremendous logistical challenge, which was solved by using a chatbot technology [45].

How Can Voice Technology Fill the Gaps?

The inability to provide in-person clinical consultations owing to the COVID-19 pandemic has fast-tracked the implementation of telehealth services. The use of virtual care solutions increased up to 10-fold within a few weeks, thus enabling patients to access clinical care remotely [46], with the majority including real-time, synchronous communication between patients and health care providers [47]. Of note, this approach is time- and resource-intensive and is thus inefficient for large patient populations. In mature telehealth systems, one telenurse can monitor up to 250 patients remotely; however, a single patient can be contacted via the telephone only periodically [48]. The COVID-19 pandemic forced an immediate implementation of new digital technologies [49,50], particularly AI-driven medical chatbots. These chatbots have the potential to improve access to health care through acute care triaging. This is favorable for COVID-19 screening and chronic disease management (long-term follow-up at home, scheduling of medical visits, and preventive care) [51]. Numerous health care systems have already utilized interactive voice response systems and chatbots to run hotlines helping to triage patients during the COVID-19 pandemic [45,52] for organizational optimization, including Massachusetts General Hospital and Brigham and Women's Hospital (Boston, Massachusetts; >40,000 digital encounters/week) [2,52], OSF Healthcare (Peoria, Illinois; >50,000 digital encounters) [53], and Providence (Seattle, Washington; >150,000 messages exchanged each day between the chatbot and users) [44].

With a user-friendly and accessible interface, voice AI chatbots provide a tool for prehospital triaging at the digital front door, assessing the clinical status of patients before they make direct contact with a health care provider. McGill University Health Centre is testing the application of a COVID-19 screening tool that uses Amazon Alexa to automatically survey patients visiting Cardiology Heart Failure Clinic (trial registration# NCT04508972) [54]. Apple's Siri provides a self-assessment tool that allows users to survey for potential COVID-19 symptoms. An automatic COVID-19 triaging service, developed on Amazon's Alexa platform, was generated by the Mayo Clinic (Rochester, Michigan) in accordance with the guidelines of the Centers for Disease Control and Prevention and fielding a substantial number of digital encounters about COVID-19 [4]. Furthermore, Apple, Amazon, and Google have removed unofficial COVID-19–related voice apps, thus preventing potential misinformation from being spread [55].

As indicated above, a mobile-responsive, web-based interface chatbot was successfully used to screen health system employees at University of California San Francisco Health, which conducted over 270,000 digital screenings within 2 months of operation. Digital solutions have optimized organizational workflow and reduced wait times for employees entering the hospital building and prevented at-risk people from coming to work [45]. OrbitaENGAGE is a voice and chat virtual assistant solution that automates critical patient engagement workflows at the so-called "digital front door" of health care. Patients interact with a voice or chatbot VA to obtain answers to health-related questions, find locations and specialists, and access symptom screening and monitoring tools for COVID-19 or other conditions including anxiety and depression [56].

Voice chatbots can potentially help patients easily communicate their health status by providing them with any disease management data. This approach allows for remote monitoring of medical patients without COVID-19 and those with COVID-19 who are mildly ill. The implementation of RPA technology integrating medical data collected through a conversational interface with the hospital database and alert-based CDSS delivers a powerful architecture that can function hand-in-hand with health care providers. Automatic clinical follow-up services provide access to up-to-date information about the individual's health status for informed medical decision-making [57] and reduce the risk of exposure and infection during face-to-face contact. More efficient patient care may help prevent unnecessary exposures due to decreased use of personal protective equipment (as exemplified by the web-based chatbot at Massachusetts General Hospital and Brigham and Women's Hospital) [2].

The implementation of innovative strategies based on VAs provides support to traditional telehealth approaches and may help reduce costs of health care services by lowering the entry bar for uninsured individuals. Direct-to-consumer digital health is a growing industry that can address unmet health care needs bypassing the traditional model (eg, that used by insurance companies), thus linking patients directly with services and providers without copays and deductibles [58]. Voice AI–supported virtual health care based on video consultations could be an alternative for people who lost employer-sponsored insurance during the COVID-19 pandemic. Virtual Care by CardioCube tests the aforementioned solution in Washington [59].

Advantages of VAs

From a user's perspective, the advancements in the field of voice-enabled technology allow for a human-like verbal communication between users and chatbots. In contrast to text-based mobile or web-based equivalents, voice chatbots have personalized speaking styles and emotions, thus providing

XSL•FO
RenderX

a more natural and intuitive experience, which is an important advantage over conventional chatbots [60,61]. Exchanging a smartphone screen with a hands-free VA might remove technological barriers. Senior users were found to prefer a conversational interface to touchscreens in a smart home environment [62]. It is also important to consider patient privacy when designing these tools. Developing multimodal solutions allows the user to choose the input modality to best suit their immediate surroundings. The user may not be comfortable speaking to VAs while in the presence of others. Having the option to provide and receive data on screen would enable them to adhere to their protocols in public environments. A strong example for the acceptance of VAs is the Healthy Coping voice bot deployed on Google Home. It is specifically targeted to patients with type 2 diabetes mellitus. The majority of users (80%) selected the voice interface over mobile solutions. Healthy Coping was assessed as easy to use and physically convenient with understandable language and communication [63]. According to Slavik et al [64], audible user interfaces are promising solutions for people with special needs for information access and control. Summarizing the advantages of voice technology, Fisher presented 5 main reasons why conversational agents may emerge as the next operating system: versatile, omnipresent, innate, contextual, and efficient characteristics (summarized as "VOICE") [65].

From a practical perspective, VAs can automate traditional telehealth services that require human providers to operate. Using conversational agents, it is possible to collect and share information at the levels of public health and individual patients. Voice chatbots can support routine care through automatic at-home monitoring, triaging, screening, providing medical recommendations and guidelines, and improving operational workflow. It is possible to help hospitals reduce their infection risk and exposure of medical staff by automatic paperless and hands-free scripting services including dictating of visit notes, charting, and patient onboarding [66].

In regard to rapid implementation, chatbot solutions are nearly off-the-shelf products that do not require substantial information technology and server infrastructure if applied with a dedicated dashboard for clinicians. The relatively low cost and rapid adoption is another important advantage of conversational agents for web-based care delivery [2].

Commercial adoption of voice technology confirms customer acceptability and provides strong grounds for the scalability and implementation of medical applications. The supportive evidence comes from the National Public Radio and Edison Research's "Smart Audio Report," which shows that there are 157 million voice devices in US households [67]. Moreover, Statista projected that the number of digital VAs in use will rise to 8 billion worldwide by 2023 [68].

Risks and Challenges

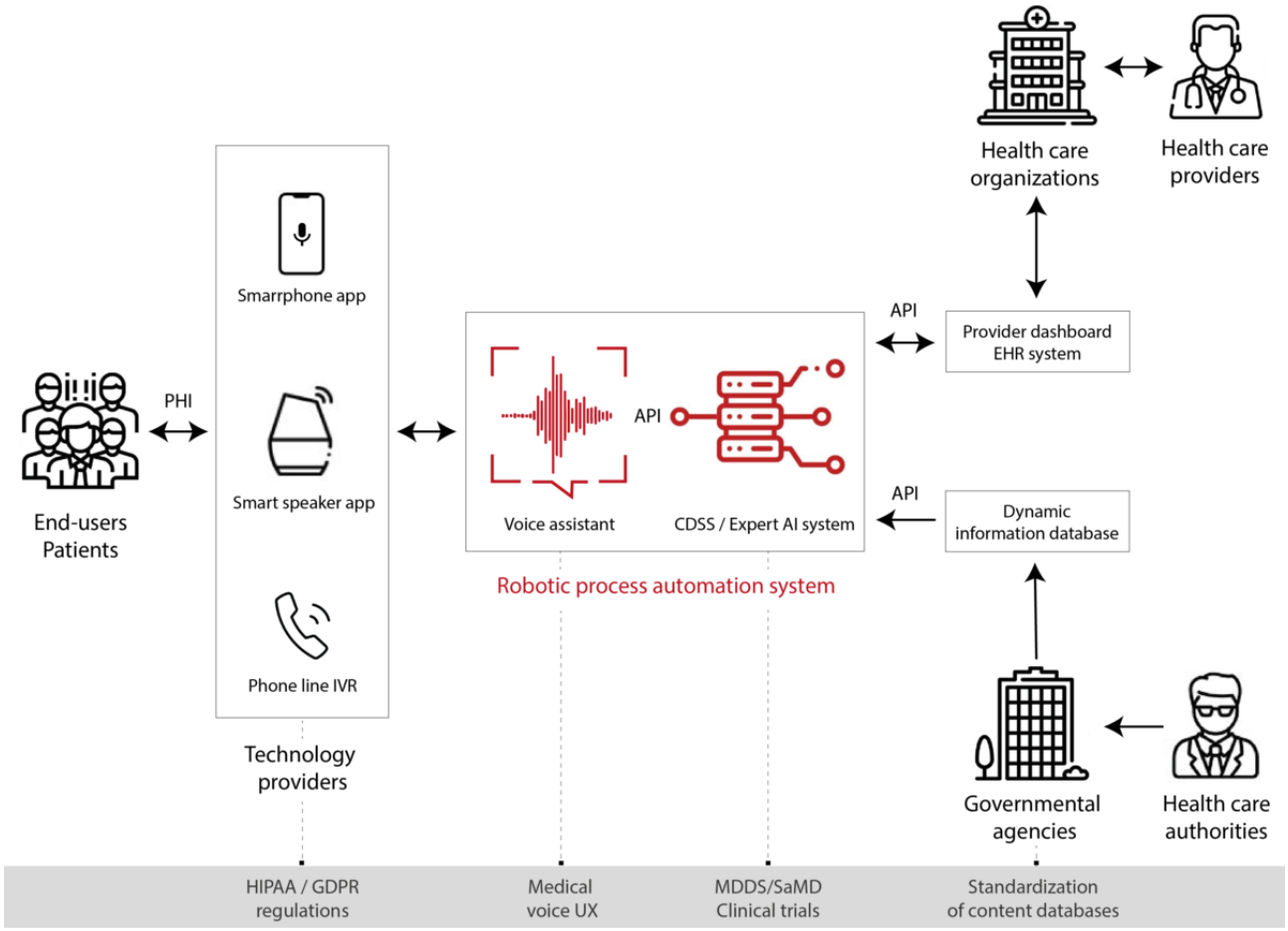
Health care systems have been working on implementing telemedicine programs for a number of years [69,70]. This leads to the question of why the uptake of web-based care solutions has been limited despite their salient clinical and economical potential. Key barriers that delay the adoption of new solutions and practice patterns exist at many levels in the chain of health care delivery [71,72]. Patient-specific obstacles include difficulties interfacing with the technology and the lack of motivation to follow through with computer-derived advice and instructions. Both patients and physicians may not wish to adopt telemedicine, as it is an imperfect surrogate for building human relationships between patients and physicians. Furthermore, both physicians and payors may be resistant to invest in web-based health platforms until these platforms have been proven to improve patient outcomes and cost-effectiveness metrics.

Initial implementation of VAs showed discordance between the quality of COVID-19–related content and guidelines by public health authorities, which have caused the dissemination of imprecise information [57]. Learning from this example, it is crucial to ensure the deliverability of reliable content through close collaboration among technology providers, developers, and health care experts. Apart from the distribution of verified information, VAs processing medical data must comply with the HIPAA to protect health information reported by end-users. Importantly, the US Department of Health and Human Services, Office for Civil Rights, waived penalties for HIPAA violation against medical providers who use everyday communication technologies to help patients remotely during the COVID-19 pandemic [39]. Furthermore, the FDA announced that the Federal Food, Drug, and Cosmetic Act will not be enforced for low-risk applications developed to combat COVID-19 [57]. These regulations supported the deployment of voice technology in clinical settings [73]. Nevertheless, it is important to establish clear regulations and guidelines for the application of voice chatbots for medical purposes. Thus far, Alexa's HIPAA-Eligible Skills program allows covered entities to design and launch HIPPA-compliant medical apps that use protected health information; these are available only in the United States [74]. However, the implementation of voice technology into clinical practice requires further regulation. In this context, guidelines and policies for telehealth (ie, those of the American Telemedicine Association) should include rules describing the utilization of medical VAs. This topic is directly associated with security, privacy, and hacking issues [75,76]. Edu et al [77] reviewed a wide range of elements that expose smart speakers at the risk of the attack surface, particularly those associated with the interaction between the user and the smart home personal assistants [77].

Amid this rapidly changing landscape, it is important to build relationships amongst the different stakeholders working to implement telemedicine innovations. These stakeholders include retail companies selling the telemedicine technology, software companies interfacing the telemedicine technology with health care systems, health care systems and their providers, payors, patients, and government health agencies (Figure 2).

XSL•FO
RenderX

Figure 2. Workflow of the AI-driven voice chatbot in health care delivery. AI: artificial intelligence; API: application programming interface; CDSS: clinical decision support systems; HER: electronic health record; GDPR: General Data Protection Regulation; HIPAA: Health Insurance Portability and Accountability Act of 1996; IVR: interactive voice response; MDDS: Medical Device Data System; PHI: protected health information; SaMD: Software as a Medical Device; UX: user experience.



Close collaboration should focus on the following:

1. Risk-based classification and verification process for VAs including assessment using FDA software as a Medical Device Guidance [15] and Medical Device Data Systems, Medical Image Storage Devices, and Medical Image Communications Devices Guidelines [14].
2. Early clinical validation; that is, using the FDA Early Feasibility Studies Program.
3. Protected health information security including voice patterns and pin-code user verification; currently implemented in HIPAA-eligible Alexa voice apps.
4. Interoperability of the systems to integrate voice-first devices, peripheral devices, and wearables with EHR systems through application programming interfaces; a feasibility study confirmed the integration of the Amazon Alexa-deployed CardioCube app with Centricity (GE Healthcare), OpenEMR, and Epic through an Fast Healthcare Interoperability Resources protocol (Cedars-Sinai Medical Center; under the Cedars-Sinai Accelerator) [78].
5. Harmonization and standardization of medical content databases providing up-to-date information through established data flow channels for quality assurance.
6. Alternative information delivery methods in case of failure in capturing or understanding questions asked by a user. Palanica et al [79] showed that VAs have different comprehension accuracy for drug names from 54.6% to 91.8% for Amazon Alexa and Google Assistant, respectively. Furthermore, Alagha and Helbing [80] showed that VAs have high variability in information quality about vaccines; hence, it is essential to secure a "safe-mode," redirecting the user to another source of information in case of delivery failure.
7. Segregation and personalization of relevant information in accordance with end-user profiles, taking into consideration determinants including demographics (especially older individuals [81,82] who showed high readiness for voice technology [83]), geographic localization, and spoken languages (voice interfaces are available in a limited number of languages) [84].
8. User experience to optimize how users who seek health information interact with VAs, preferably dedicated medical voice user designers with expertise in health care to understand patients' needs.
9. Context-aware method of delivery dependent on end-user characteristics and the digital ecosystem; for example, voice-only as a stand-alone smart speaker, smartphone- or computer-deployed app (more applicable for users with broadband internet access), or the telephone (useful for rural areas with limited internet access) vs voice and video devices (ie, Google Nest). In this aspect, VAs might be

XSL•FO
RenderX

used for provider-patient audio calls, while devices with embedded video cameras allow audio-visual communication.

Reimbursement for VAs is an important factor determining its adoption. Even though chatbots are not currently eligible for billing both in Europe and in the United States, Keesara et al [50] proposed that the payment structure could be based on time-based models or fixed fee-for-service monetization. Furthermore, the reimbursement for evaluation and management billing codes could be adjusted for digital services, while the Centers for Medicare and Medicaid Services could remove restrictions on in-person consultations under evaluation and management services [50]. Alternatively, having well-established distribution, supply chain, and AI services, software companies developing VAs could become a key player in the health care ecosystem. Interestingly, Amazon Care provides app-based health services to its employees bypassing health plans and brokers. Furthermore, Amazon has enabled patients to order and have their medications delivered home through Amazon Pharmacy. Subsequently, with Amazon's Pillo Health, an interactive drug dispenser with voice-first technology, patients can optimize pharmacological treatment at home [85]. In the case of health care organizations, Amazon released HealthLake, a HIPAA-eligible data management service functioning as a cloud EHR system [86].

Future Directions

VAs could be applied to capture human voice utilized as a digital biomarker. The analysis of vocal characteristics is a promising research field that combines AI and clinical medicine. Maor et al [87] reported an independent association between voice signal analysis and hospitalization as well as mortality among patients with heart failure. Furthermore, quantitative voice analysis was shown to be applicable in the diagnosis of neurodegenerative diseases [88] and COVID-19 [89].

Conclusions

In summary, with the growing demand for telehealth services, VAs could extend the workforce of medical care providers by using AI-powered interfaces to ensure the safety of patients and medical staff. Conversational agents have the potential to become a regular component of health care systems, thereby multiplying medical capacities during the current COVID-19 pandemic and reducing the spread of COVID-19. Moreover, clinical-grade voice AI chatbots can sustainably supplement routine clinical work in the postpandemic period. Voice technology implements synergistic and practical solutions, which have the potential to optimize health care systems and increase preparedness for future COVID-19–like pandemics. Currently, the Amazon Alexa HIPAA-eligible environment enables VAs implemented in the United States to be applied as a professional medical tool.

Acknowledgments

The authors would like to thank Dr Teri Fisher, physician and clinical assistant professor at the University of British Columbia; Nathan Treloar, co-founder, president, and chief operating officer at Orbita Inc; and David Box, global managing director, Digital Healthcare and Wellness at Star for insightful comments and stimulating discussions. This study was funded by statutory funds from the Medical University of Silesia (KNW-1-135/N/6/K), ReGenHeart (H2020, grant# 731532), SCIENCE (H2020, grant# 643478), and Circulate (STRATEGMED II, grant# Strategmed2/265761/10/NCBR/2015) trials.

Conflicts of Interest

TJ is the co-founder and chief science officer at CardioCube Corp (Seattle, Washington). WW is a shareholder in CardioCube Corp. The other authors declare no conflict of interest.

Multimedia Appendix 1
Video recording presenting CardioCube VA for patients with diabetes.
[MOV File , 115187 KB - jmir_v23i5e22959_app1.mov]

References

1. Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Med Ref Serv Q 2018;37(1):81-88. [doi: 10.1080/02763869.2018.1404391] [Medline: 29327988]
2. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: Novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020 Dec;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]
3. Sezgin E, Noritz G, Elek A, Conkol K, Rust S, Bailey M, et al. Capturing At-Home Health and Care Information for Children With Medical Complexity Using Voice Interactive Technologies: Multi-Stakeholder Viewpoint. J Med Internet Res 2020 Feb 13;22(2):e14202 [FREE Full text] [doi: 10.2196/14202] [Medline: 32053114]
4. Furst J. Mayo Clinic introduces skill for Amazon's Alexa about COVID-19. Mayo Clinic. 2020 Apr 27. URL: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-introduces-skill-for-amazons-alexa-about-covid-19/ [accessed 2020-03-21]

XSL•FO
RenderX

5. Beaman J. Interactive Voice-Based Administration of the PHQ-9. ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04609267 [accessed 2021-05-17]
6. Palanica A, Thommandram A, Lee A, Li M, Fossat Y. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. NPJ Digit Med 2019;2:55 [FREE Full text] [doi: 10.1038/s41746-019-0133-x] [Medline: 31304401]
7. Jadczyk T, Kiwic O, Khandwalla RM, Grabowski K, Rudawski S, Magaczewski P, et al. Feasibility of a voice-enabled automated platform for medical data collection: CardioCube. Int J Med Inform 2019 Sep;129:388-393. [doi: 10.1016/j.ijmedinf.2019.07.001] [Medline: 31445282]
8. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
9. Nobles AL, Leas EC, Caputi TL, Zhu S, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. NPJ Digit Med 2020;3:11 [FREE Full text] [doi: 10.1038/s41746-019-0215-9] [Medline: 32025572]
10. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of Conversational Agents to Health and Lifestyle Prompts: Investigation of Appropriateness and Presentation Structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
11. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075 [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
12. Koulaouzidis G, Charisopoulou D, Wojakowski W, Koulaouzidis A, Marlicz W, Jadczyk T. Telemedicine in cardiology in the time of coronavirus disease 2019: a friend that everybody needs. Pol Arch Intern Med 2020 Jun 25;130(6):559-561 [FREE Full text] [doi: 10.20452/pamw.15432] [Medline: 32539313]
13. Medical Device Data Systems. US Food and Drug Administration. 2019 Sep 26. URL: https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/medical-device-data-systems [accessed 2021-03-21]
14. Medical Device Data Systems, Medical Image Storage Devices, and Medical Image Communications Devices. US Food and Drug Administration. 2019. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/medical-device-data-systems-medical-image-storage-devices-and-medical-image-communications-devices [accessed 2021-03-21]
15. Software as a Medical Device (SaMD). US Food and Drug Adminisration. 2018 Apr 12. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd [accessed 2021-03-21]
16. WebMD. Amazon. URL: https://www.amazon.com/gp/product/B01MRM361G [accessed 2021-03-21]
17. Cleveland Clinic Tip of the Day. Amazon. URL: https://www.amazon.com/Cleveland-Clinic-Tip-Day/dp/B077BFFQSZ [accessed 2021-03-21]
18. KidsMD. Amazon. URL: https://www.amazon.com/Boston-Childrens-Hospital-KidsMD/dp/B01E3F0AUC [accessed 2021-03-21]
19. Mayo Clinic First Aid. Amazon. URL: https://www.amazon.com/mayo-clinic-first-aid/dp/b0744ljcv2 [accessed 2021-03-21]
20. First Aid. Amazon. URL: https://www.amazon.com/American-Red-Cross-First-Aid/dp/B07KMSZQ5F [accessed 2021-03-21]
21. Mayo Clinic News Network. Amazon. URL: https://www.amazon.com/mayo-clinic-news-network/dp/b0711hkbb6 [accessed 2021-03-21]
22. MyChildren's Enhanced Recovery After Surgery. Amazon. URL: https://www.amazon.com/MyChildrens-Enhanced-Recovery-After-Surgery/dp/B07QB7PQYW/ [accessed 2021-03-21]
23. OhioHealth: Alexa skills. Amazon. URL: https://www.amazon.com/OhioHealth/dp/B077SNT8L6 [accessed 2021-03-21]
24. New Hanover Regional Medical Center. Amazon. URL: https://www.amazon.com/New-Hanover-Regional-Medical-Center/dp/B07BS5TGSQ [accessed 2021-03-21]
25. Livongo Blood Sugar Lookup. Amazon. URL: https://www.amazon.com/Livongo-Health-Blood-Sugar-Lookup/dp/B07QHF76RN [accessed 2021-03-21]
26. Giant Eagle Pharmacy. Amazon. URL: https://www.amazon.com/Omnicell-Inc-Giant-Eagle-Pharmacy/dp/B08164DYJ7 [accessed 2021-03-21]
27. Express Scripts. Amazon. URL: https://www.amazon.com/gp/product/B07QB7P6Y2 [accessed 2021-03-21]
28. Swedish Health Connect: Alexa skills. Amazon. URL: https://www.amazon.com/gp/product/B07PGJYYF6 [accessed 2021-03-21]
29. Atrium Health. Amazon. URL: https://www.amazon.com/gp/product/B07H1T46DC, [accessed 2021-03-21]
30. OrbitaASSIST: Virtual Bedside Assistant. Orbita. URL: https://orbita.ai/orbita-assist-virtual-bedside-assistant/ [accessed 2021-03-21]
31. Answers by Cigna. Amazon. URL: https://www.amazon.com/Cigna-Answers-by/dp/B079SMKCZB [accessed 2021-03-21]
32. OrbitaCONNECT. Orbita. URL: https://orbita.ai/orbita-connect-virtual-health-assistant-care-continuity/, [accessed 2021-03-21]

XSL•FO
RenderX

33. Talkspace: Online therapy for all. Amazon. URL: https://www.amazon.com/Talkspace-Online-therapy-for-all/dp/B085VHRW33 [accessed 2021-03-21]
34. Personalized Voice Activated Wellness Assistants for Patients With Heart Failure. ClinicalTrials.gov. URL: https://www.clinicaltrials.gov/ct2/show/NCT03707275 [accessed 2021-03-21]
35. Haase CB, Bearman M, Brodersen J, Hoeyer K, Risor T. 'You should see a doctor', said the robot: Reflections on a digital diagnostic device in a pandemic age. Scand J Public Health 2021 Feb 18;49(1):33-36 [FREE Full text] [doi: 10.1177/1403494820980268] [Medline: 33339468]
36. Iansiti M, Lakhani K. Competing in the Age of AI. Harvard Business Review. 2020 Jan 16. URL: https://hbr.org/webinar/2020/01/competing-in-the-age-of-ai#:~:text=and%20deliver%20value.-,In%20their%20new%20book%20Competing%20in%20the%20Age%20of%20AI,scale%2C%20scope%2C%20and%20learning [accessed 2021-05-17]
37. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558 [FREE Full text] [doi: 10.1016/S1473-3099(20)30144-4] [Medline: 32171059]
38. Schwirtz M. Nurses die, doctors fall sick and panic rises on virus front lines. The New York Times. URL: https://www.nytimes.com/2020/03/30/nyregion/ny-coronavirus-doctors-sick.html [accessed 2021-03-21]
39. Rose C. Am I Part of the Cure or Am I Part of the Disease? Keeping Coronavirus Out When a Doctor Comes Home. N Engl J Med 2020 Apr 30;382(18):1684-1685. [doi: 10.1056/NEJMp2004768] [Medline: 32187461]
40. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med 2020 Apr;26(4):506-510 [FREE Full text] [doi: 10.1038/s41591-020-0822-7] [Medline: 32284616]
41. Worldometer. URL: https://www.worldometers.info/coronavirus/ [accessed 2021-03-21]
42. Solomon MD, McNulty EJ, Rana JS, Leong TK, Lee C, Sung S, et al. The Covid-19 Pandemic and the Incidence of Acute Myocardial Infarction. N Engl J Med 2020 Aug 13;383(7):691-693. [doi: 10.1056/NEJMc2015630] [Medline: 32427432]
43. Krouse HJ. COVID-19 and the Widening Gap in Health Inequity. Otolaryngol Head Neck Surg 2020 Jul;163(1):65-66. [doi: 10.1177/0194599820926463] [Medline: 32366172]
44. Lacina L. COVID-19 reveals gaps in health systems: WHO Briefing. World Economic Forum. 2020 May 06. URL: https://www.weforum.org/agenda/2020/05/covid-19-reveals-gaps-in-public-health-system-who-briefing/ [accessed 2021-03-21]
45. Judson T, Odisho A, Young J, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
46. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
47. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
48. Cosentino D. Ten steps to building a successful telehealth program. Caring 2009 Jul;28(7):34-39. [Medline: 19772085]
49. Schulman KA, Richman BD. Toward an Effective Innovation Agenda. N Engl J Med 2019 Mar 07;380(10):900-901. [doi: 10.1056/NEJMp1812460] [Medline: 30673357]
50. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
51. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
52. Schwamm LH, Erskine A, Licurse A. A digital embrace to blunt the curve of COVID19 pandemic. NPJ Digit Med 2020;3:64 [FREE Full text] [doi: 10.1038/s41746-020-0279-6] [Medline: 32377575]
53. Roth M. How OSF'S 6 Digital Approaches Handle More Than 50,000 COVID Encounters. HealthLeaders. 2020 Apr 28. URL: https://www.healthleadersmedia.com/innovation/how-osfs-6-digital-approaches-handle-more-50000-covid-encounters [accessed 2021-03-21]
54. The VOICE-COVID-19. ClinicalTrials.gov. URL: https://www.clinicaltrials.gov/ct2/show/NCT04508972 [accessed 2021-03-21]
55. Sherr I. Apple, Google, Amazon block nonofficial coronavirus apps from app stores. CNET. 2020 Mar 15. URL: https://www.cnet.com/news/apple-google-amazon-block-nonofficial-coronavirus-apps-from-app-stores/ [accessed 2021-03-21]
56. Employee Health Manager. Orbita. URL: https://orbita.ai/orbita-employee-health-manager/ [accessed 2021-03-21]
57. Ross C. I asked eight chatbots whether I had Covid-19. The answers ranged from ‘low’ risk to ‘start home isolation’. STAT. 2020 Mar 23. URL: https://www.statnews.com/2020/03/23/coronavirus-i-asked-eight-chatbots-whether-i-had-covid-19/ [accessed 2021-03-21]
58. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health 2020 Apr;2(4):e163-e165 [FREE Full text] [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]

XSL•FO
RenderX

59. Virtual healthcare at an affordable rate. CardioCube. URL: https://cardiocube.com/ [accessed 2021-03-21]
60. Gao C. Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience. Amazon Alexa. 2019. URL: https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2019/11/new-alexa-emotions-and-speaking-styles [accessed 2021-03-21]
61. Gao C. Create Engaging Voice Experiences with a New Alexa Speaking Style, Speaking Styles in Amazon Polly Voices and Additional Polly Voices. Amazon Alexa. 2020. URL: https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2020/04/new-alexa-long-form-speaking-style-and-polly-voices?tag=theverge02-20 [accessed 2021-03-21]
62. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2011 Oct 2;17(1):127-144. [doi: 10.1007/s00779-011-0470-5]
63. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. 2018 Presented at: 2018 15th IEEE Annual Consumer Communications & Networking Conference; January 12-15, 2018; Las Vegas, NV p. 12-15. [doi: 10.1109/ccnc.2018.8319283]
64. Slavík P, Němec V, Sporka AJ. Speech Based User Interface for Users with Special Needs. 2005 Presented at: 8th International Conference on Text, Speech and Dialogue; September 12-15, 2005; Karlovy Vary. [doi: 10.1007/11551874_6]
65. Fisher T, Metcalf D, Pruthi S, Pappas H. The Opportunity for Voice in Healthcare. In: Voice Technology in Healthcare: Leveraging Voice to Enhance Patient and Provider Experiences. New York, NY: Productivity Press; 2020.
66. Kumah-Crystal Y, Pirtle C, Whyte H, Goode E, Anders S, Lehmann C. Electronic Health Record Interactions through Voice: A Review. Appl Clin Inform 2018 Jul;9(3):541-552 [FREE Full text] [doi: 10.1055/s-0038-1666844] [Medline: 30040113]
67. The Smart Audio Report. National Public Media. 2020. URL: https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/, [accessed 2021-03-21]
68. Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)*. Statista. URL: https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/#statisticContainer, [accessed 2021-03-21]
69. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
70. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: A review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
71. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
72. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
73. FCNcare A new way to connect with your care team through Alexa. FNCare. URL: https://fcncare.com/ [accessed 2021-03-21]
74. Voice for Healthcare. Amazon Alexa. URL: https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills/healthcare-skills [accessed 2021-03-21]
75. Lau J, Zimmerman B, Schaub F. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. Proc ACM Hum-Comput Interact 2018 Nov;2:1-31. [doi: 10.1145/3274371]
76. Trebilcock O. Are Alexa and Google Assistant spying on us? Which?. 2020 Jan 24. URL: https://www.which.co.uk/news/2020/01/are-alexa-and-google-assistant-spying-on-us/ [accessed 2021-03-21]
77. Edu JS, Such JM, Suarez-Tangil G. Smart Home Personal Assistants. ACM Comput Surv 2021 Feb;53(6):1-36. [doi: 10.1145/3412383]
78. The Cedars-Sinai Accelarator. Cedars Sinai. URL: https://csaccelerator.com/ [accessed 2021-03-21]
79. Palanica A, Thommandram A, Lee A, Li M, Fossat Y. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. NPJ Digit Med 2019;2:55 [FREE Full text] [doi: 10.1038/s41746-019-0133-x] [Medline: 31304401]
80. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075 [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
81. Reis A, Paulino D, Paredes H. Using intelligent personal assistants to assist the elderlies An evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri. 2018 Presented at: 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing; June 20-22, 2018; Thessaloniki. [doi: 10.1109/tishw.2018.8559503]
82. Trajkova M, Martin-Hammond A. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. 2020 Presented at: CHI Conference on Human Factors in Computing Systems; April 2020; Honolulu, HI. [doi: 10.1145/3313831.3376760]
83. Kowalska M, Gładyś A, Kalańska-Łukasik B, Gruz-Kwapisz M, Wojakowski W, Jadczyk T. Readiness for Voice Technology in Patients With Cardiovascular Diseases: Cross-Sectional Study. J Med Internet Res 2020 Dec 17;22(12):e20456 [FREE Full text] [doi: 10.2196/20456] [Medline: 33331824]

XSL•FO
RenderX

84. Language Support in Voice Assistants Compared. Summa Linguae Technologies. 2021 Apr 21. URL: https://www.globalme.net/blog/language-support-voice-assistants-compared/ [accessed 2021-03-21]
85. Jercich K. Amazon plans to provide health services to workers at other companies, says BI. Healthcare IT News Australia. 2020 Dec 16. URL: https://www.healthcareitnews.com/news/amazon-plans-provide-health-services-workers-other-companies-says-bi [accessed 2021-03-21]
86. Amazon HealthLake (Preview): Store, transform, query, and analyze health data in minutes. Amazon Web Services Inc. URL: https://aws.amazon.com/healthlake/ [accessed 2021-03-21]
87. Maor E, Perry D, Mevorach D, Taiblum N, Luz Y, Mazin I, et al. Vocal Biomarker Is Associated With Hospitalization and Mortality Among Heart Failure Patients. J Am Heart Assoc 2020 Apr 07;9(7):e013359 [FREE Full text] [doi: 10.1161/JAHA.119.013359] [Medline: 32233754]
88. Mekyska J, Galaz Z, Kiska T, Zvoncak V, Mucha J, Smekal Z, et al. Quantitative Analysis of Relationship Between Hypokinetic Dysarthria and the Freezing of Gait in Parkinson's Disease. Cognit Comput 2018;10(6):1006-1018 [FREE Full text] [doi: 10.1007/s12559-018-9575-8] [Medline: 30595758]
89. Han J, Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Spathis D, et al. Exploring Automatic COVID-19 Diagnosis via Voice and Symptoms from Crowdsourced Data. 2021 Presented at: 2021 IEEE International Conference on Acoustics, Speech and Signal Processing; June 6-11, 2021; Toronto, ON. [doi: 10.1109/ICASSP39728.2021.9414576]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support systems
EHR: electronic health record
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act of 1996
RPA: robotic process automation
SaMD: Software as a Medical Device
VA: voice assistant

Edited by C Basch; submitted 28.07.20; peer-reviewed by T Jamieson, Y Fossat, P Yellowlees, I Gabashvili; comments to author 02.09.20; revised version received 20.02.21; accepted 21.03.21; published 25.05.21.

Please cite as:
Jadczyk T, Wojakowski W, Tendera M, Henry TD, Egnaczyk G, Shreenivas S
Artificial Intelligence Can Improve Patient Management at the Time of a Pandemic: The Role of Voice Technology
J Med Internet Res 2021;23(5):e22959
URL: https://www.jmir.org/2021/5/e22959
doi:10.2196/22959
PMID:33999834

©Tomasz Jadczyk, Wojciech Wojakowski, Michal Tendera, Timothy D Henry, Gregory Egnaczyk, Satya Shreenivas. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Improving the Usability and Safety of Digital Health Systems: The Role of Predictive Human-Computer Interaction Modeling

Chris Paton[1,2], BMedSci, BMBS, MBA; Andre W Kushniruk[3], BSc, BA, MSc, PhD; Elizabeth M Borycki[3], RN, HBScN, MN, PhD; Mike English[1], MBBChir, MD, FMedSci; Jim Warren[4], BS, PhD

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[2]Department of Information Science, University of Otago, Dunedin, New Zealand
[3]School of Health Information Science, University of Victoria, Victoria, BC, Canada
[4]School of Computer Science, University of Auckland, Auckland, New Zealand

Corresponding Author:
Chris Paton, BMedSci, BMBS, MBA
Nuffield Department of Medicine
University of Oxford
Peter Medawar Building
South Parks Road
Oxford, OX1 3SY
United Kingdom
Phone: 44 7552 698058
Email: chris.paton@ndm.ox.ac.uk

Abstract

In this paper, we describe techniques for predictive modeling of human-computer interaction (HCI) and discuss how they could be used in the development and evaluation of user interfaces for digital health systems such as electronic health record systems. Predictive HCI modeling has the potential to improve the generalizability of usability evaluations of digital health interventions beyond specific contexts, especially when integrated with models of distributed cognition and higher-level sociotechnical frameworks. Evidence generated from building and testing HCI models of the user interface (UI) components for different types of digital health interventions could be valuable for informing evidence-based UI design guidelines to support the development of safer and more effective UIs for digital health interventions.

(J Med Internet Res 2021;23(5):e25281) doi:10.2196/25281

KEYWORDS

digital health; human-centered design; usability; human-computer interaction; predictive modeling

Introduction

User interfaces for digital health systems such as electronic health records (EHRs) or clinical decision support systems should be designed so that clinicians can accomplish tasks efficiently without making errors that could compromise patient safety. Designers of digital health systems should be able to use the best research evidence currently available, drawn from systematic reviews and meta-analyses, to inform their designs. However, the evidence base for designing user interfaces (UIs) of digital health systems has been difficult to establish, as evaluations of UIs are often subjective and difficult to generalize to new clinical contexts [1]. Recent systematic reviews of usability issues with different types of digital health systems (such as computerized physician order entry [2] and electronic medical records [3]) highlight some common issues identified across different studies but also describe the difficulties in generalizing guidance from context-specific evaluations. This evidence is of use to designers but does not offer specific design patterns or quantitatively demonstrate the trade-offs between efficiency and effectiveness that may be involved in different approaches to making designs more usable. Partly due to the weakness of the scientific evidence base, usability guidelines have therefore generally recommended adopting a human-centered design (HCD) approach and the use of expert heuristics to guide the design of interfaces rather than quantitatively validated design patterns.

In this paper, we examine how the use of human-computer interaction (HCI) predictive models can contribute to building a more robust and generalizable evidence base for UI designs for digital health interventions. This evidence base could then be used to advance UI design guidelines and could be

XSL•FO
RenderX

incorporated in the human-centered design process to accelerate innovation and improve clinical safety.

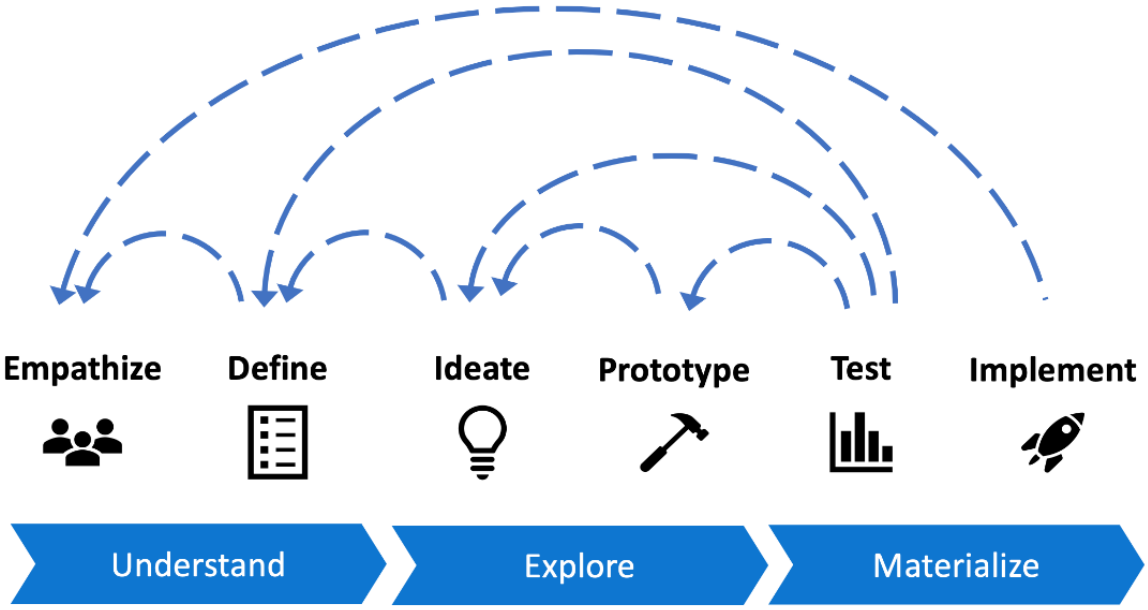
HCI modeling was used to develop the first computer mouse [4,5] and the modern window-based graphical UIs in wide use today [6]. As digital health systems become widely (albeit often reluctantly) adopted in health care, HCI modeling could have an important role in ensuring that the systems we use to care for patients are as safe and effective as other medical innovations such as drugs and diagnostics.

To discuss this approach, we provide a summary of the historical origins of predictive HCI modeling, show examples of how it can be used in modern digital health system design, and show how HCI modeling can be integrated into the human-centered design process.

The Digital Health Design Evidence Gap

The current best practice for designing UIs for digital health systems is to use a human-centered design (HCD) approach such as "design thinking" [7], in which designers and developers move through an iterative and flexible process of understanding, exploring, and materializing the end product (Figure 1). This process is often guided by design heuristics ("rules of thumb") that include such guidance as keeping the UI simple and aesthetically pleasing and ensuring that help and documentation are readily available (see Textbox 1 for the 10 Nielsen heuristics) [8,9,10]. Iterative design thinking methods attempt to ensure that systems are aligned with users' behaviors and needs and allow for improvements to be made throughout the course of the design process.

Figure 1. Human-centered design helps designers move from computer code to real-world use. Adapted from Gibbons [7].



Textbox 1. The 10 Nielsen usability heuristics.

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

HCD has been developed to ensure that UIs work well for specific contexts but does not provide the kind of evidence normally expected for medical interventions. HCD should be part of the design process; however, additional methods are

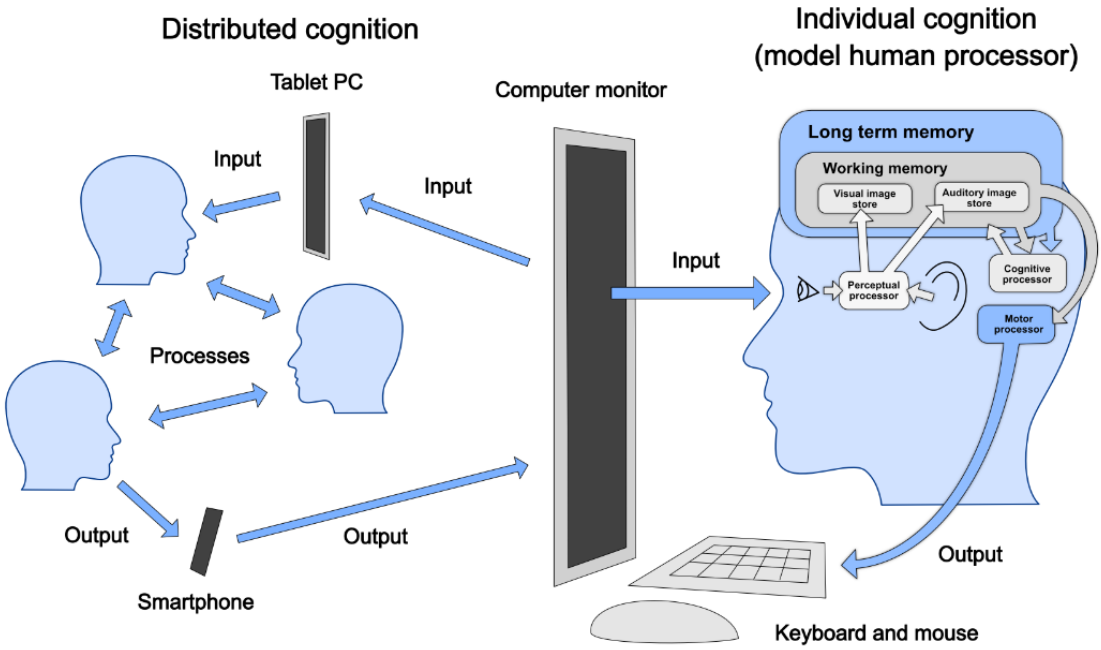
XSL•FO
RenderX

needed with a more scientific basis to be confident that digital health UI designs are suitable for use in high-risk settings such as hospitals.

Predictive Models of HCI

Predictive HCI models have the potential to explain how users interact with digital health interventions at the level of individual human cognition. For more than 50 years, empirically derived predictive HCI models have been used for ensuring the safety and usability of information systems (physical and digital) for industrial and commercial applications ranging from avionics to power plant control systems. These models informed the designs of the first desktop computers, with innovations such as the computer mouse and windows-based graphical UIs. The "classic" HCI model was called the "model human processor," in which the different components of human cognitive systems were modeled and combined with models of the interactions (inputs and outputs) between the human and the computer system (see Figure 2).

Figure 2. The model human processor: a model of how long it takes to process information (from perception to action) and how we can use the limited "chunks" of information in working memory. Building on the idea of a model human processor is the concept of "distributed cognition," with multiple humans and devices working together. Adapted from Card et al [11].



There are now several types of HCI predictive models that can be used depending on the context or type of tasks analyzed. For example, the Fitts law is used for mouse pointing [12], the Hick-Hyman law is used for reviewing a sorted list [13,14], and goals operators methods and selection rules (GOMS) [15] is a high-level model that is used to describe the cognitive processes and methods involved in using a computer to achieve specific goals. The keystroke-level model (KLM) is a more specific type of GOMS model that is used for compositions of tasks that fit how an experienced user interacts with the interface [11]. KLM has also recently been updated for touch interfaces [16]. Although individuals will vary in their performance speed, these models give a good indication of the relative effort required to accomplish a task. For example, Warren et al and others [17-19] have used such models for simulation-based evaluation of split-menu designs (placing commonly accessed options at the top of a list) for clinical information systems. These models are primarily aimed at reducing the time needed to complete tasks by eliminating unnecessary clicks and ensuring that UI elements are easy to navigate. In the health care domain, simplifying designs and increasing efficiency is also likely to reduce errors that can cause patient harm, such as choosing incorrect items from unnecessarily long menus or clicking through lengthy screens too quickly [20,21].

Limitations of HCI Modeling

Although the abovementioned models proved useful and effective in the design of early graphical UIs and input devices, the ways in which teams of people began to use computer systems in the 1990s prompted a move in HCI research communities away from the micro-level interactions to meso-level systems of "distributed cognition" [22] (Figure 2) and "situated action" [23]. These systems included the described cognitive models by attempting to place them within a social context, with human-human interaction playing a role in addition to machine-machine interaction (Figure 2 shows how distributed cognition can be integrated with micro-level HCI modeling). In the health care domain, Borycki and Kushniruk [20,24-26] led the development of an integrative cognitive-sociotechnical model for characterizing user interactions with health care systems at multiple levels.

The field of cognitive science has also moved on since HCI modeling was first proposed. The human cognitive system

XSL•FO
RenderX

processes information in a highly complex manner, with dynamic feedback loops that are not included in the ways that HCI models are traditionally described. Therefore, HCI modeling has been more recently viewed as a useful tool for assisting in the development of systems rather than the scientific pursuit originally envisaged by Card, Moran, and Newell in their classic text "The Psychology of Human-Computer Interaction" [15]. However, recent work, such as on the representation of "M" (for "mentally prepare") in the KLM model, shows that more nuanced and complex models can be created where more fidelity to human cognition is required.

Example of the "Combined" Layered Approach to Collecting Evidence About HCI

In their study of electronic whiteboards across Denmark, Rasmussen and Kushniruk [27] incorporated the KLM model; however, this work was initially driven by a naturalistic study to locate workflow issues associated with deployment of a new electronic whiteboard across several hospitals across Denmark. The naturalistic study identified areas where optimization might be needed, as evidenced by observation of seemingly inefficient user interactions with the whiteboard (identified from review of the screen recordings of user interactions), and the KLM experimental approach was then used to test hypotheses on applying different changes to create efficiencies.

The research team video recorded 2863 entries from video analysis of live user interactions with the whiteboard and identified potential inefficient sequences from observing and timing the video (eg, the task "add new patient" took an average of 12.3 seconds). They then conducted GOMS-KLM analysis producing the following predictive model (H=move hands; M=mentally prepare; K=tap key or button; P=point):

H+M+P+K+M+P+K+H+M+ (K*10) +H+M+P+K+P+K+H+M+ (K*30)+H+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K+M+P+K = 54.6 seconds

They then modified the "add new patients" task so that instead of opening new dialog boxes, information could be directly entered into text boxes or from menus, modeled thus:

H+M+P+K+M+P+K+H+M+K+ (K*10) +H+M+P+K+H+M+ (K*30) +H+M+P+K+M+P+K+M+P+K+M+P+K=31.25 seconds

The results of the project led to a reduction in time to task completion of 44.6% and illustrated the benefits of considering HCI at multiple levels (ie, including the use of naturalistic observation and video coding of those data for the use of KLM to predict more optimal user designs for improving inefficient user interaction sequences).

HCI Model Patterns for Different Types of Digital Health Systems

Digital health applications may have common usability and resultant patient safety issues that can be modeled using predictive HCI approaches as shown above. However, specific types of digital health interventions may also have type-specific UI patterns that, if modeled as a common function of a particular type of system, may make it easier to develop more general models. Using the World Health Organization Digital Health Intervention (DHI) classification system [28], it could be envisaged that each type of system, such as a telemedicine system (DHI 2.4) or health care provider training system (DHI 2.8), would have a common HCI predictive model that takes into account the cognitive processes involved in using that type of system. For example, a training system would include cognitive models of how the system enables the user to learn how to manage a clinical problem, retain the knowledge over time (perhaps by repeatedly "topping up" their knowledge), and recall the information when needed.

Implications for Developing Guidelines and Standards for Digital Health Systems

Safety issues with large-scale EHR systems have now started to be reported in the literature. Recent work by Ratwani et al [29,30] has highlighted a wide range of usability issues in currently used digital health systems. The analysis by Pacheco et al [31] of the Certified Health IT Product List database of usability attestation, for example, revealed that 3.7% of the products surveyed had a certified capability nonconformity issue that was coded as being associated with possible patient harm. Despite these ongoing issues with clinical usability, current guidance on digital health system development has largely taken a heuristic approach [32], building on and adapting the Nielsen heuristics for health care contexts [3]. Several large-scale projects have been undertaken to develop UI guidelines; however, without establishing an evidence-based approach to guideline development, it has been difficult to maintain or build on these guidelines as technology develops. For example, in the United Kingdom, the National Health Service commissioned Microsoft to develop a "Common User Interface" for EHR systems designed as part of the National Programme for Information Technology; however, this guidance was recently withdrawn without replacement [33]. In the United States, the Department of Health and Human Services created "Research-Based Web Design and Usability Guidelines" [34].

Although these approaches can aid the design of systems that adhere to industrial usability standards, they represent a broad-brush approach that lacks the kind of scientific rigor required by other health care interventions, such as new pharmaceuticals.

Greater consideration and use of predictive models integrated into an HCD approach may be needed to ensure that evidence-based UI design guidelines can be developed over time. The results of modeling-based studies could ensure that systematic reviews and meta-analyses of usability studies generate evidence that is generalizable beyond the specific contexts of the studies.

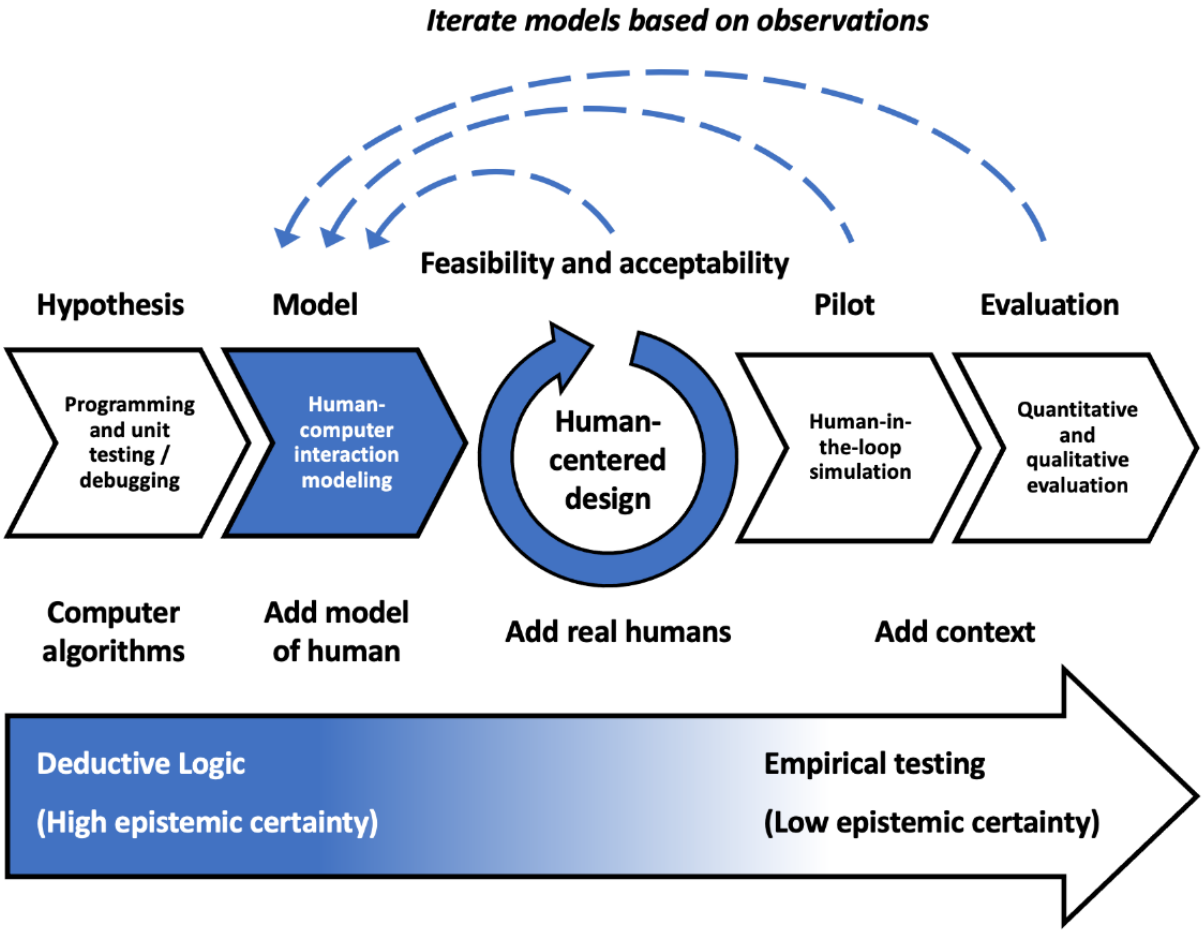
Integrating Predictive Modeling With HCD

Using predictive models to inform the HCD process could accelerate the design of digital health systems. By having a

XSL•FO
RenderX

validated evidence base of UI designs to draw on, designers could eliminate a large number of potential designs that might meet basic usability heuristics or that could be appealing to early testers but that could be shown through predictive modeling to have poor usability. Figure 3 shows how HCI modeling could fit into the process of designing the UI for a digital health application. The design process moves from implementing the computer algorithms needed for the software to function (developed using deductive logic with a high level of epistemic certainty) to modeling how users would interact with the UI of the system using HCI cognitive models. Once designs that show poor results with modeling are weeded out, the project will then enter a human-centered design phase in which the system is trialed with real users (for example, nurses and physicians who will use a digital health system on the wards) and repeatedly iterated until the software is sufficiently acceptable to pilot. At this stage, human-in-the-loop simulations can be conducted as the system is piloted. Finally, more formal quantitative and qualitative evaluations in clinical contexts can provide higher-level empirical evidence (albeit with lower epistemic certainty than with in-silico HCI modeling). At each stage, in keeping with the design thinking approach, the development team can move back to modeling and HCD to improve the design if needed. In addition to showing whether a particular application works, real-world evaluations based on HCI models will show which models work in the real world, building the evidence base for future design guidelines.

Figure 3. Predictive human-computer interaction modeling could augment the human-centered design process and help us understand how an application achieves real-world effectiveness.



Conclusion

UIs for digital health applications are currently designed using techniques developed for commercial software applications based on human-centered design processes and heuristics. Predictive HCI modeling of applications may help improve the design process and allow for more scientific progress toward safer and more effective digital health systems. We have described in this paper how predictive HCI modeling has developed from individual cognitive modeling to distributed cognitive models and provided examples of how these models can be integrated into sociotechnical modeling approaches. Although predictive HCI modeling has fallen out of favor in recent years, as the demand for more evidence of the safety and effectiveness of digital health systems increases, it is worth re-evaluating whether HCI modeling can contribute to the science of evidence-based digital health system design. Future research on the integration of predictive modeling with usability and software engineering approaches (such as usability testing and human-in-the-loop simulations) is both needed and warranted.

XSL•FO
RenderX

Acknowledgments

This viewpoint paper resulted from a panel discussion session by the authors on the topic of HCI modeling at the International Medical Informatics Association MedInfo Conference held in Lyon, France, in 2019. We are grateful to the attendees of the session for their helpful comments and questions during the panel discussion.

This research was funded by the National Institute for Health Research (NIHR; NIHR130812: Learning to Harness Innovation in Global Health for Quality Care (HIGH-Q)) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. Rigby M, Ammenwerth E. The Need for Evidence in Health Informatics. Stud Health Technol Inform 2016;222:3-13. [Medline: 27198087]
2. Khajouei R, Jaspers MWM. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: a systematic review. Methods Inf Med 2010;49(1):3-19. [doi: 10.3414/ME0630] [Medline: 19582333]
3. Zahabi M, Kaber DB, Swangnetr M. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. Hum Factors 2015 Aug;57(5):805-834. [doi: 10.1177/0018720815576827] [Medline: 25850118]
4. Card SK, English WK, Burr BJ. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. Ergonomics 2007 Apr 25;21(8):601-613. [doi: 10.1080/00140137808931762]
5. English W, Engelbart D, Berman M. Display-Selection Techniques for Text Manipulation. IEEE Trans. Human Factors Electron 1967 Mar;HFE-8(1):5-15. [doi: 10.1109/thfe.1967.232994]
6. Scott MI. Human-Computer Interaction: An Empirical Research Perspective. Burlington, Massachusetts, United States: Morgan Kaufmann; 2013.
7. Gibbons S. Design Thinking 101. Nielsen Norman Group. URL: https://www.nngroup.com/articles/design-thinking/ [accessed 2020-06-23]
8. Nielsen N. 10 usability heuristics for user interface design. Norman Nielsen Group. 2020. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2021-05-12]
9. Nielsen J. Enhancing the explanatory power of usability heuristics. : ACM; 1994 Presented at: CHI94: SIGCHI Conference on Human Factors in Computing Systems; 1994; Boston Massachusetts USA. [doi: 10.1145/191666.191729]
10. Nielsen J. Heuristics for user interface design. Nielsen Norman Group. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2020-06-23]
11. Card SK, Moran TP, Newell A. The keystroke-level model for user performance time with interactive systems. Commun. ACM 1980 Jul;23(7):396-410. [doi: 10.1145/358886.358895]
12. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 1954;47(6):381-391. [doi: 10.1037/h0055392]
13. Hick WE. On the Rate of Gain of Information. Quarterly Journal of Experimental Psychology 2018 Jan 01;4(1):11-26. [doi: 10.1080/17470215208416600]
14. Hyman R. Stimulus information as a determinant of reaction time. Journal of Experimental Psychology 1953;45(3):188-196. [doi: 10.1037/h0056940]
15. Card S, Newell A, Moran T. The Psychology of Human-Computer Interaction. Boca Raton, Florida, United States: CRC Press; Feb 01, 1986.
16. Rice A, Lartigue J. Touch-level model (TLM). 2014 Presented at: ACM SE '14: Proceedings of the 2014 ACM Southeast Regional Conference; 2014; Kennesaw Georgia p. 1-6. [doi: 10.1145/2638404.2638532]
17. Canfield K. Priming intelligent split menus with text corpora for computerized patient record data-entry. International Journal of Bio-Medical Computing 1995 May;39(2):263-273. [doi: 10.1016/0020-7101(95)01108-Q]
18. Warren JR, Bolton P. Intelligent split menus for data entry: a simulation study in general practice medicine. Proc AMIA Symp 1999:450-454 [FREE Full text] [Medline: 10566399]
19. Sears A, Shneiderman B. Split menus. ACM Trans. Comput.-Hum. Interact 1994 Mar;1(1):27-51. [doi: 10.1145/174630.174632]
20. Borycki E, Kushniruk A, Nohr C, Takeda H, Kuwata S, Carvalho C, et al. Usability Methods for Ensuring Health Information Technology Safety: Evidence-Based Approaches. Contribution of the IMIA Working Group Health Informatics for Patient Safety. Yearb Med Inform 2013;8:20-27. [Medline: 23974544]

XSL•FO
RenderX

21. Marcilly R, Peute L, Beuscart-Zephir M. From Usability Engineering to Evidence-based Usability in Health IT. Stud Health Technol Inform 2016;222:126-138. [doi: 10.3233/978-1-61499-635-4-126]
22. Hollan J, Hutchins E, Kirsh D. Distributed cognition. ACM Trans. Comput.-Hum. Interact 2000 Jun;7(2):174-196. [doi: 10.1145/353485.353487]
23. Bardram J. Plans as Situated Action: An Activity Theory Approach to Workflow Systems. Dordrecht: Springer; 1997 Presented at: Fifth European Conference on Computer Supported Cooperative Work; 1997; Netherlands. [doi: 10.1007/978-94-015-7372-6_2]
24. Kushniruk A, Nohr C, Borycki E. Human Factors for More Usable and Safer Health Information Technology: Where Are We Now and Where do We Go from Here? Yearb Med Inform 2016(1):120-125. [doi: 10.15265/iy-2016-024]
25. Borycki E, Kushniruk A, Keay E, Nicoll J, Anderson J, Anderson M. Toward an Integrated Simulation Approach for Predicting and Preventing Technology-Induced Errors in Healthcare: Implications for Healthcare Decision-Makers. hcq 2009 Aug 15;12(sp):90-96. [doi: 10.12927/hcq.2009.20974]
26. Kushniruk A, Patel V. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
27. Rasmussen R, Kushniruk A. The long and twisting path: an efficiency evaluation of an electronic whiteboard system. Stud Health Technol Inform 2013;183:174-178. [Medline: 23388277]
28. World HO. Classification of Digital Health Interventions.: World Health Organization; 2018. URL: https://apps.who.int/iris/handle/10665/260480 [accessed 2020-02-19]
29. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic Health Record Usability Issues and Potential Contribution to Patient Harm. JAMA 2018 Mar 27;319(12):1276. [doi: 10.1001/jama.2018.1171]
30. Ratwani RM, Savage E, Will A, Fong A, Karavite D, Muthu N, et al. Identifying Electronic Health Record Usability And Safety Challenges In Pediatric Settings. Health Affairs 2018 Nov;37(11):1752-1759. [doi: 10.1377/hlthaff.2018.0699]
31. Pacheco TB, Hettinger AZ, Ratwani RM. Identifying Potential Patient Safety Issues From the Federal Electronic Health Record Surveillance Program. JAMA 2019 Dec 17;322(23):2339. [doi: 10.1001/jama.2019.17242]
32. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
33. Edgar M. Why we are deprecating and then withdrawing the NHS Common User Interface - NHS Digital. NHS Digital. 2019 Jul 04. URL: https://digital.nhs.uk/blog/transformation-blog/2019/why-we-are-withdrawing-the-nhs-common-user-interface [accessed 2019-12-18]
34. Shneiderman B, Leavitt M, Others. Research-based web design & usability guidelines. Washington DC: Department of Health and Human Services; 2006.

Abbreviations

DHI: digital health intervention
EHR: electronic health record
GOMS: goals operators methods and selection rules
HCD: human-centered design
HCI: human-computer interaction
KLM: keystroke-level model
UI: user interface

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by A Hilt, R Marcilly, E Andrikopoulou, D Lin, J Ye; comments to author 23.12.20; revised version received 17.02.21; accepted 16.03.21; published 27.05.21.

Please cite as:
Paton C, Kushniruk AW, Borycki EM, English M, Warren J
Improving the Usability and Safety of Digital Health Systems: The Role of Predictive Human-Computer Interaction Modeling
J Med Internet Res 2021;23(5):e25281
URL: https://www.jmir.org/2021/5/e25281
doi:10.2196/25281
PMID:34042590

©Chris Paton, Andre W Kushniruk, Elizabeth M Borycki, Mike English, Jim Warren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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,

XSL•FO
RenderX

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

Tutorial

Determination of Patient Sentiment and Emotion in Ophthalmology: Infoveillance Tutorial on Web-Based Health Forum Discussions

Anne Xuan-Lan Nguyen[1]; Xuan-Vi Trinh[2]; Sophia Y Wang[3*], MD; Albert Y Wu[3*], MD, PhD, FACS

[1]Faculty of Medicine, McGill University, Montreal, QC, Canada

[2]Department of Computer Science, McGill University, Montreal, QC, Canada

[3]Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, CA, United States

[*]these authors contributed equally

Corresponding Author:
Albert Y Wu, MD, PhD, FACS
Department of Ophthalmology
Byers Eye Institute
Stanford University
2452 Watson Court
Palo Alto, CA, 94303
United States
Phone: 1 650 497 0758
Email: awu1@stanford.edu

Abstract

Background: Clinical data in social media are an underused source of information with great potential to allow for a deeper understanding of patient values, attitudes, and preferences.

Objective: This tutorial aims to describe a novel, robust, and modular method for the sentiment analysis and emotion detection of free text from web-based forums and the factors to consider during its application.

Methods: We mined the discussion and user information of all posts containing search terms related to a medical subspecialty (oculoplastics) from MedHelp, the largest web-based platform for patient health forums. We used data cleaning and processing tools to define the relevant subset of results and prepare them for sentiment analysis. We executed sentiment and emotion analyses by using IBM Watson Natural Language Understanding to generate sentiment and emotion scores for the posts and their associated keywords. The keywords were aggregated using natural language processing tools.

Results: Overall, 39 oculoplastic-related search terms resulted in 46,381 eligible posts within 14,329 threads. Posts were written by 18,319 users (117 doctors; 18,202 patients) and included 201,611 associated keywords. Keywords that occurred ≥500 times in the corpus were used to identify the most prominent topics, including specific symptoms, medication, and complications. The sentiment and emotion scores of these keywords and eligible posts were analyzed to provide concrete examples of the potential of this methodology to allow for a better understanding of patients' attitudes. The overall sentiment score reflects a positive, neutral, or negative sentiment, whereas the emotion scores (anger, disgust, fear, joy, and sadness) represent the likelihood of the presence of the emotion. In keyword grouping analyses, medical signs, symptoms, and diseases had the lowest overall sentiment scores (−0.598). Complications were highly associated with sadness (0.485). Forum posts mentioning body parts were related to sadness (0.416) and fear (0.321). Administration was the category with the highest anger score (0.146). The top 6 forum subgroups had an overall negative sentiment score; the most negative one was the *Neurology* forum, with a score of −0.438. The *Undiagnosed Symptoms* forum had the highest sadness score (0.448). The least likely fearful posts were those from the *Eye Care* forum, with a score of 0.260. The overall sentiment score was much more negative before the doctor replied. The anger, disgust, fear, and sadness emotion scores decreased in likelihood, whereas joy was slightly more likely to be expressed after doctors replied.

Conclusions: This report allows physicians and researchers to efficiently mine and perform sentiment analysis on social media to better understand patients' perspectives and promote patient-centric care. Important factors to be considered during its application include evaluating the scope of the search; selecting search terms and understanding their linguistic usages; and establishing selection, filtering, and processing criteria for posts and keywords tailored to the desired results.

(J Med Internet Res 2021;23(5):e20803) doi:10.2196/20803

XSL•FO
RenderX

KEYWORDS

sentiment analysis; emotions analysis; natural language processing; online forums; social media; patient attitudes; medicine; infodemiology; infoveillance; digital health

Introduction

Understanding patient attitudes and expectations toward health care is an important component of promoting patient-centric care and patient satisfaction. However, studies have shown that physicians have difficulties in understanding patients' health beliefs and concerns [1]. Strategies to improve the understanding of patient attitudes have traditionally required the development of specialized survey instruments, which may nonetheless be limited in scope, or focus groups, which can be very time consuming and laborious [2].

The internet has now become a rich additional source of information regarding patients' attitudes and expectations toward health care. Recent decades have seen a rapid increase in internet engagement, with an estimated 5 billion people using mobile devices [3], and more than half of the global population actively using the internet [4]. In 2012, 72% of American internet users sought health information on the web [5] and many also increasingly expressed their medical concerns on the web [6,7]. These web-based communication outlets include social networks (eg, Facebook, Twitter, or Instagram), doctor review websites (eg, Healthgrades, Vitals, or RateMDs), and health web forums (eg, MedHelp, Health245, or Patient info). Analyzing people's health-related queries and reports on the internet to better inform public health and public policy is an increasingly popular field known as infoveillance [8]. Although Twitter is a common and popular platform based on which many infoveillance studies are conducted, its space-limited format contrasts with web-based health forums, which are a particularly rich resource for understanding patient attitudes toward medical issues by supporting patients in directly seeking medical advice, sharing their medical experiences, and discussing their symptoms at length [9-15].

Understanding unstructured clinical data on social media requires natural language processing (NLP), a well-established branch of artificial intelligence that has been applied in a variety of fields and has emerging applications in medicine [16,17]. Sentiment and emotion analyses, which are subbranches of NLP, can identify and quantify positive, neutral, and negative sentiments and can detect emotions such as anger, disgust, fear, joy, and sadness in free text [18,19]. The data mining and sentiment analysis of social media, especially web-based medical discussion forums, can provide a fast and effective way to better understand patients' attitudes, expectations, and experiences [18], which can better guide patient-centric care [20]. The literature shows that health care professionals can, with the sentiment analysis of web-based medical forums, discover new outlooks of patient issues and recurrent complications related to specific treatment uses and drugs [19,21,22] and administrative burden and access to care [23]. By analyzing forum posts, physicians can further understand patients' attitudes and experiences and assess their needs and concerns, which can result in better patient-centric care [24].

We examined all oculoplastics-related posts on MedHelp, which included questions from patients and replies written by patients and doctors. Oculoplastics is a subspecialty in ophthalmology that involves the eyelids, face, tear ducts, and orbit and is both highly specialized and interdisciplinary as a clinical domain, often at the intersection of ophthalmology, plastic surgery, dermatology, and otolaryngology. Our study illustrates the challenges of identifying and distinguishing text related to specialized medical subdomains, such as ophthalmology, in the context of patient-centric idiomatic language and of web-based discussion forum analysis, where the relevance of text must be filtered on multiple structural levels and physician and patient posts must be distinguished from physicians' posts. We provide all scripts and describe a detailed approach toward web-based patient forum sentiment analysis, which includes data collection; rigorous data processing, cleaning, and selection; and in-depth data analysis. This methodology allows for a variety of applications, notably the identification and analysis of the main topics related to the chosen field (eg, symptoms, complications, and medication) and their associated quantified sentiment (positive, neutral, or negative) and the likelihood of the presence of certain emotions (joy, anger, disgust, sadness, and fear). This methodology can also be used as a means to measure patient satisfaction and perspective by comparing patients' sentiment and emotions before, during, and after their interaction with health care professionals. This paper aims to guide physicians and researchers to mine and perform sentiment analysis on web-based clinical data in a chosen field and highlights the challenges and approaches to consider in the process.

Methods

Data Source and Study Population

Founded in 1994, MedHelp is the world's largest web-based health community [25]. With more than 15 million visits per month, it allows users (patients and doctors) to discuss issues related to various health and wellness topics on a daily basis [18]. Currently, this platform contains 299 official support communities, including a wide variety of well-established medical discussion forums. The main oculoplastic discussion forum is the *Eye Care Community*, which encourages patients to discuss eye-related issues. Another vision-related forum was the *Ask a Doctor-Eye Care Forum*, which benefited from a collaboration with ophthalmologists from the American Academy of Ophthalmology from 2007 to 2014 [25,26]. In addition to these forums, MedHelp has more than 1000 user-made groups.

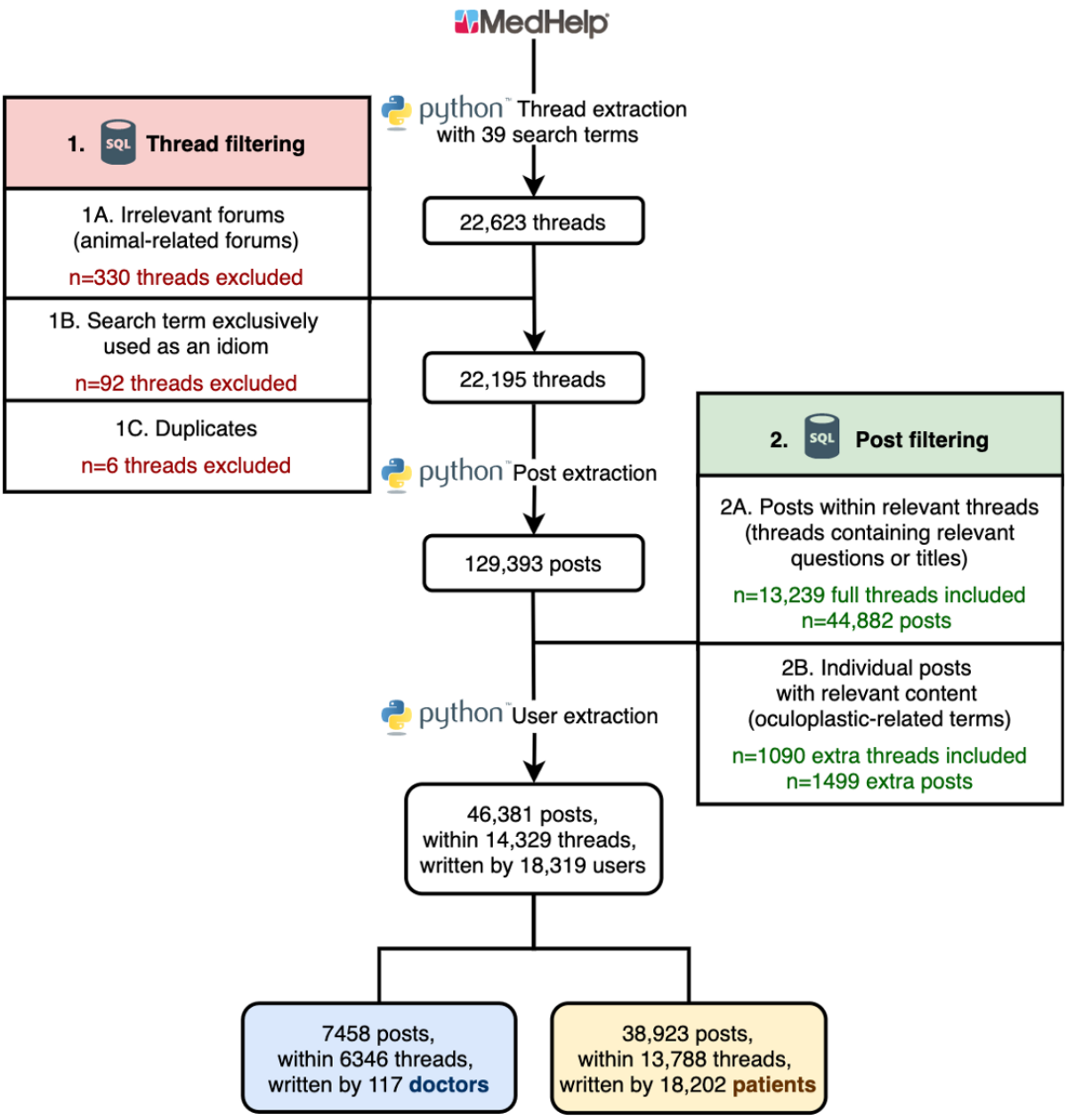
Each community or group, also referred to as *a forum*, encompasses various discussion *threads*. Discussion threads comprise a question asked by a user (the initial *post*), followed by replies written by individual users, which are also considered *posts* [19].

XSL•FO
RenderX

Approach to Data Extraction

The approach to data extraction from MedHelp is summarized in Figure 1. Discussion threads related to oculoplastic surgery were identified from MedHelp using a list of oculoplastics-relevant search terms created by consensus between 2 specialized ophthalmologists, AYW and SYW, and AXN (Multimedia Appendix 1).

Figure 1. Flowchart for the data extraction of discussion threads and posts on web-based medical forums. SQL: Structured Query Language.



Each discussion thread was parsed using a Python script (Python Software Foundation, version 3.8.6) [27] and the Python package *Beautiful Soup* [28] to yield the full text of each post (including the initial question and all replies) and the relevant metadata, including the MedHelp user for each post and the forum that each thread belonged to.

An initial review of the search results demonstrated that not all results appeared to be relevant, and it was noted that the details of the exact algorithm used by MedHelp's proprietary search engine could not be known. Thus, we performed additional filtering of the search results to remove irrelevant discussion threads. Threads in animal forums, duplicate threads, and threads where the search terms were mentioned in purely idiomatic ways were removed.

In addition, we noted that many threads were returned as search results because search terms appeared in different posts within the same thread, for example, the search term "double eye lid" could return a thread containing the use of "double," "eye," and "lid" in separate posts, which could result in many irrelevant posts.

Therefore, to further filter the posts to include those that were most highly relevant to oculoplastics, we developed additional lists of related terms and text patterns and identified all the posts that contained exact matches to these patterns (Multimedia

XSL•FO
RenderX

Appendices 2-3) after lowercasing all the posts. Posts were deemed relevant and included for analyses if they were (1) in a thread whose title or initial question contained an exact pattern match (Multimedia Appendix 2) or (2) the post itself contained an exact pattern match to a very specific oculoplastics-related term (Multimedia Appendix 3). Posts that were not part of a relevant thread were subject to more stringent inclusion criteria because the original topic of the thread did not necessarily pertain to oculoplastics. This filtering algorithm ensures that the data set is relevant and tailored and is not influenced by the proprietary search algorithm of the platform.

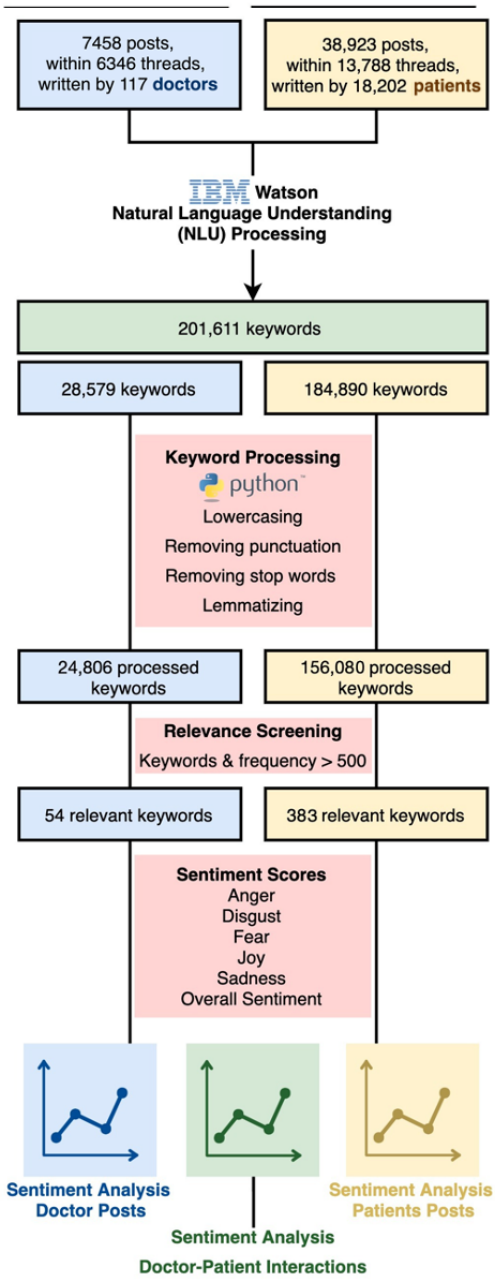
Patterns required for inclusion of posts allowed for some variability in human language, for example, the two patterns "%upper lid%eye" and "eye%upper lid%" ("%" denotes 0 or more of any character) match a subset of posts expressing one's upper eyelid, such as "my eye hurts, and my upper lid..." and "my upper lid droops, and my eye keeps twitching," without deeming posts containing solely "upper lid" as relevant, such as "the upper lid of my jar...." After excluding irrelevant posts, we extracted the username, user type (doctor or patient), self-reported age, sex, registration date to the MedHelp community, and user location from each user profile. All data were stored in an SQLite relational database [29]. The scripts used to extract threads, posts, and users and the detailed instructions on how to use them can be found in our repository [30].

Approach to Natural Language Understanding Processing

The approach to NLP and sentiment analysis is presented in Figure 2 [31]. We used IBM's Watson Natural Language Understanding (NLU; IBM Cloud Natural Language Understanding V1, version 2019-07-12) [32] to perform sentiment and emotion analyses on the free text of every included forum post. The Watson machine learning system reads and understands the semantics of free text by breaking down sentences structurally, grammatically, and contextually through various linguistic models and algorithms. The results that were returned included a sentiment score for the full document (ie, the full text of a single post) and for each keyword extracted by the IBM Watson algorithm and emotion scores for anger, sadness, joy, fear, and disgust at both the post and keyword levels. These keywords include important words, entities, and phrases from each post. Sentiment scores ranged from −1 to +1 on an arbitrary linear scale of intensity and were negative (less than 0), neutral (0), or positive (greater than 0). For each emotion, a score was given in the form of a percentage of likelihood, ranging from 0 to 1, where 0 represents the certain absence of the emotion in question and 1 represents the definite presence of the emotion.

XSL•FO
RenderX

Figure 2. Flowchart describing keyword processing and sentiment analysis.



NLU Keywords Processing

Related keywords generated by the IBM Watson NLU program were processed using a Jupyter Notebook [33] with Natural Language Toolkit (NLTK) [34], NumPy [35], and Pandas [36] libraries. The following transformations were applied to each keyword: lowercasing, punctuation removal, stop word deletion (eg, prepositions and conjunctions), and lemmatization [37] (morphological destructuring that allows words to be stripped down to their root word, eg, “oculoplastics” into “oculoplastic”).

NLU Keywords Selection and Categorization

Among the keywords with a frequency higher than 500, manual verification was performed to merge the keywords with the same semantic meaning. These keywords were then classified into various categories (groups and subgroups). For example, the “people” group encompasses multiple subgroups including the “eye care provider” subgroup, which in turn contains the fully processed keywords “ophthalmologist” and “optometrist.” However, keywords with a questionable relevancy to the clinical field and keywords with a general meaning (eg, “thing,” “thought,” and “name”) were excluded from the analysis.

Sentiment Scores Statistical Analysis

We used Python to aggregate and calculate the mean and standard deviation of each keyword’s associated sentiment and emotion scores (sentiment, sadness, fear, anger, joy, and disgust scores). Three examples of the analyses were performed with the results. We performed a summary of statistics by keyword grouping to determine significant trends among the chosen clinical categories. We also analyzed the data by forum subgroups (eg, posts in the *Eye Care* forum vs posts in the *Neurology* forum). We also compared the sentiment associated

XSL•FO
RenderX

with the posts written by the patient before a doctor replied with the patient's posts written after a doctor replied.

Results

Results From Data Extraction

Threads Extraction and Filtering

Searching the 300 forums (including ongoing communities, discontinued forums, and user-made groups) on MedHelp using 39 oculoplastics-related search terms resulted in 22,623 discussion threads (Multimedia Appendix 1). The screening for irrelevant threads resulted in the exclusion of 6 duplicate threads, 330 threads found in animal-related forums, and 92 threads containing the search term used exclusively as an idiom. Table 1 highlights threads containing the common idioms associated with the initial search term lists and excluded forums (*Animal Health—General*, *Animal Lovers Group*, *Animal-Surgery*, *Birds*, *Cats*, *Dogs*), as well as example text from the excluded threads and the associated number of threads deleted.

Table 1. Examples of excluded posts because of idiomatic language or reference to animals.

Idiom or forum name	Description	Threads deleted, n (%)	Example text from excluded threads
Idiom			
(1) Raise an *eyebrow*[a]; (2) raise an *eye brow*; (3) raise *eyebrows*	This idiom is used to convey awe, consternation, or disbelief.	(1) 51 (100); (2) 2 (100); (3) 18 (64)	"I may be just freaking out but it does raise an *eyebrow.*"
(1) Bat an *eyelid*; (2) bat an *eye lid*	This idiom is used to show an emotional reaction.	(1) 20 (100); (2) 1 (100)	"And the doctor, like me, has seen so many she's not going to bat an *eyelid*!"
Forum			
Animal Health—General	This forum is used to answer questions related to general pet health (treatment, parasites, infectious disease, etc).	56 (100)	"My 3 year old boxer has one *eye* that seems to droop and is a little redder than normal. [...] It has always been that way it could be a congenital abnormality such as *entropion*."
Animal Lovers Group	This forum was previously used to chat about anything related to pets and animals.	1 (100)	"Birds are wonderful. In this state, they seem to *frown* on folks feeding them in the park too, it really irritates me, what would our world be like without those lovely creatures singing their happy song to us, I love them."
Animal-Surgery	This forum was previously used to have questions answered by a veterinarian from PetDocsOnCall on all questions regarding animal surgery.	2 (100)	"My dog has ingrown *eyelashes*"
Birds	This forum was used to answer questions about pet birds.[b]	5 (100)	"My three year old peacock has cloudy *eyes*. One *eye* in particular, the lid seems to linger and appears to bulge (slightly) when looking at him straight."
Cats	This forum was used to answer questions about pet cats.[b]	113 (100)	"I don't know what my cat has got into but his left *eye* has been watering really bad and is red inside. It is now red on the right *eye* but just around where the *lashes* would be."
Dogs	This forum was used to answer questions about pet dogs.[b]	153 (100)	"Lumps on dogs *eye lid*"

[a]Words referring to ophthalmology are italicized.

[b]These forums used to have questions answered by a veterinarian.

Posts Extraction and Filtering

After filtering the threads, 129,393 posts associated with the resulting 22,195 threads remained, which then underwent additional layers of filtering for inclusion and exclusion (Figure 1). Posts from 13,239 of the 22,195 threads were considered relevant and were therefore included because the thread title or question contained a relevant oculoplastic term (Multimedia Appendix 2), which resulted in 44,882 included posts. An additional 1499 individual posts from 1090 other discussion threads also contained oculoplastic-related keywords (Multimedia Appendix 3) and were therefore included in the analysis. The final corpus was composed of 46,381 posts within 14,329 threads, which were written between January 1, 1995, and December 18, 2019, in 273 forums.

User Extraction

These 46,381 posts were written by 18,319 users from 1995 to 2019. More specifically, 7458 posts (within 6346 threads) were written by 117 doctors, and 38,923 posts (within 13,788 threads) were written by 18,202 patients. Overall, 20.19% (3699/18,319) of users were male patients, 38.33% (7022/18,319) were female patients, 40.84% (7481/18,319) of the patients did not specify their sex, 0.41% (75/18,319) were male doctors, and 0.23%

XSL•FO
RenderX

(42/18,319) were female doctors. A total of 5642 patients were included in this study. Their ages varied from 10 to 96 years, with an average of 44.8 years. A total of 6704 patients indicated their location (city, state, and/or country).

Results From Keyword Processing

Keyword Extraction

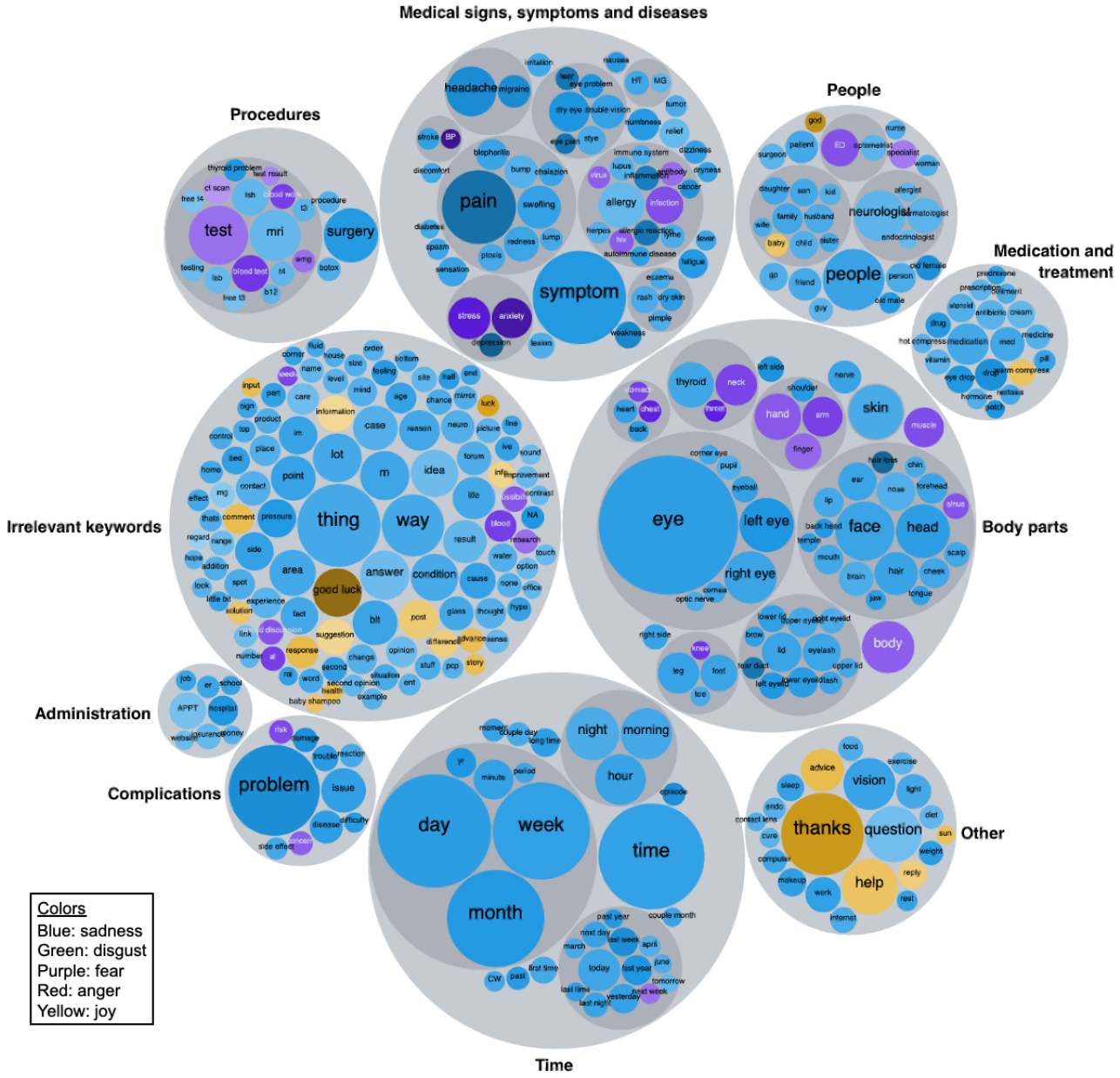
Keyword extraction, sentiment analysis, and emotion analysis were performed using the IBM Watson NLU service, which generated 201,611 unique raw keywords, including 28,579 keywords from posts written by doctors and 184,890 keywords from posts written by patients, with some keywords common to both sets of posts (Figure 2). Further processing using the NLTK Python library grouped related keywords, resulting in 24,806 keywords from doctors' posts and 156,080 keywords from patients' posts. For instance, "eyes" became "eye," "eyelids" became "eyelid," and "eye lashes" became "eye lash."

Keyword Selection and Categorization

Keywords that occurred at least 500 times in the corpus were included for analysis; 383 keywords were from patients' posts and 54 keywords were from doctors' posts. We grouped these keywords into nine relevant categories: body parts; medical signs, symptoms, and diseases; people; medication and treatment; procedures; complications; administration; aggravating and relieving factors; and others. Some of these categories were then subdivided into more precise clinical concepts. For example, the broad category *body parts* contained keywords related to the head, neck, upper limbs, thorax, and lower limbs. The category *medical signs, symptoms, and diseases* was subdivided by specialty (oculoplastics, ophthalmology, psychiatry, neurology, endocrinology, integumentary, immunology, cardiology, and gastroenterology). The *people* category contained references to eye care doctors, nonocular medical specialists, surgeons, family doctors, family members, friends, and other health care professionals (Figure 3) [38].

XSL•FO
RenderX

Figure 3. Nested bubble chart showing the top 500 keywords associated with patient posts and grouped into clinically relevant categories. The size of each bubble is proportional to the frequency of the keyword. The color of each bubble represents the most likely emotion associated with the keyword. The shade of each bubble is proportional to the likelihood of the emotion score; emotions that are more likely are in darker bubbles. APPT: appointment; BP: blood pressure; ED: eye doctor; HP: hypothyroidism; MG: myasthenia gravis.



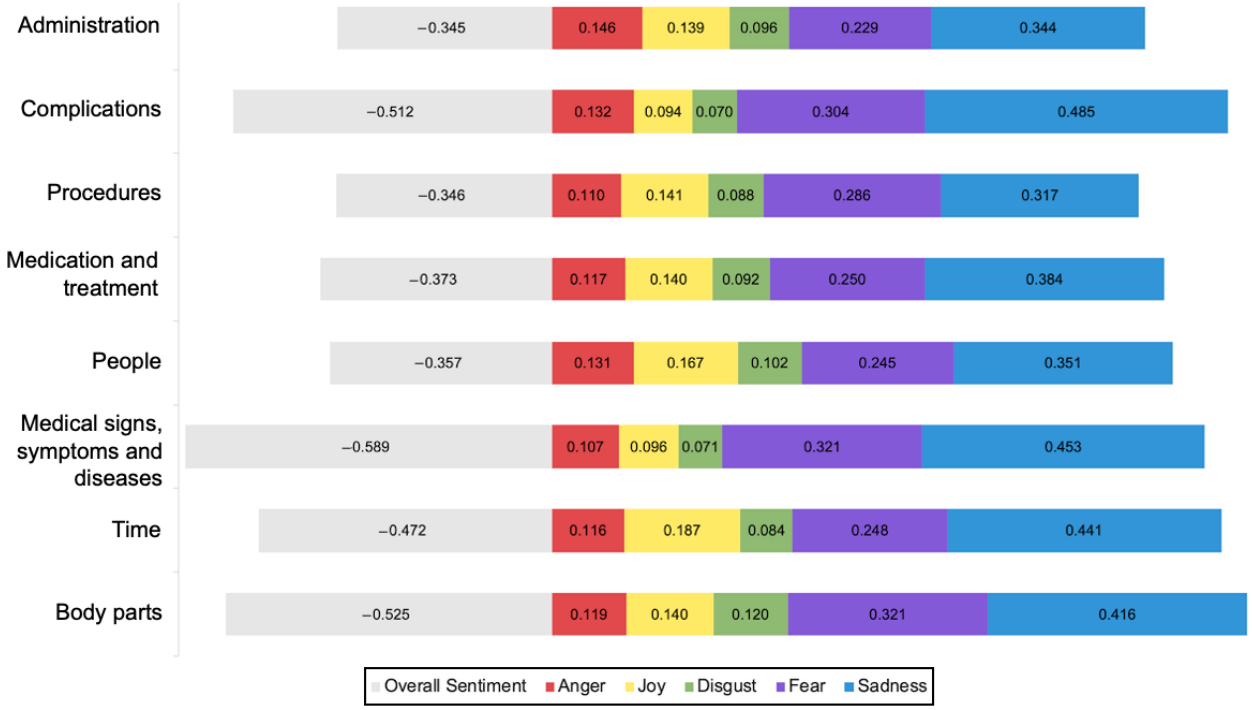
Similar keywords that were aggregated include the following examples: “itch” encompassing both “itch” and “itching,” “diagnosis” replacing “dx” and “diagnosis,” “eyelid” including both “eyelid” and “eye lid,” “eyebrow” (“eye brow” and “eyebrow”), “twitch” (“twitch” and “twitching”), “treatment” (“tx” and “treatment”), “non specified doctor” (“doctor,” “doc,” “dr,” “physician,” and “md”), and “ophthalmologist” (“ophthalmologist” and “ophthamologists” [sic]).

Sentiment and Emotion Analysis

Summary statistics were therefore performed using keyword groupings (Figures 3 and 4). Medical signs, symptoms, and diseases had the lowest overall sentiment scores (−0.598). Complications were highly associated with sadness (likelihood sadness score of 0.485). Forum posts mentioning body parts were related to sadness (likelihood sadness score of 0.416) and fear (likelihood fear score of 0.321). Administration was the category with the highest anger score (0.146).

XSL•FO
RenderX

Figure 4. Top 8 groupings and their respective overall sentiment and emotion scores. The overall sentiment score reflects a positive, neutral, or negative sentiment, whereas the emotion score (anger, disgust, fear, joy, and sadness) represents how likely (%) the emotion is to be present.



We further analyzed sentiments and emotions by the forum subgroup. We compared the most popular forums among each other by analyzing the sentiment and emotion scores of their posts (Multimedia Appendix 4). All 6 forums had an overall negative sentiment score; the most negative one being the *Neurology* forum with a score of −0.438. The *Undiagnosed Symptoms* forum had the highest sadness score (0.448). The least likely fearful posts were those from the *Eye Care* forum, with a score of 0.260.

We also analyzed all the posts from users who asked questions (ie, initiated new threads) on MedHelp. These posts were divided into two categories: the pre–doctor reply group and the post–doctor reply group. The pre–doctor reply group included all the questions, the self-replies, and replies to other users written by the initial user before a doctor replied. The post–doctor reply group included all the other posts written by the initial user after the first doctor replied. As seen in Table 2, the overall sentiment score is much more negative before the doctor replied. We can also see shifts in the emotion scores: anger, disgust, fear, and sadness decreased in likelihood whereas joy was expressed slightly more likely after the doctor replied.

Table 2. Difference in sentiment and emotion scores between the posts written before and after a doctor replied.

Posts analyzed	Pre–doctor reply group	Post–doctor reply group	Difference (post − pre)
Posts expressing the following sentiment			
Negative, n (%)	1553 (92.22)	1260 (49.55)	−42.67%
Neutral, n (%)	11 (0.65)	110 (4.33)	+3.67%
Positive, n (%)	120 (7.13)	1172 (46.09)	+38.97%
Scores			
Overall sentiment	−0.557	0.0268	+0.584
Anger	0.143	0.109	−0.0334
Disgust	0.126	0.0740	−0.0505
Fear	0.364	0.233	−0.130
Joy	0.308	0.348	+0.0391
Sadness	0.5210	0.335	−0.186

XSL•FO
RenderX

Discussion

Innovation

This is the first paper providing a detailed methodology for preparing unstructured text data from web-based health discussion forums related to ophthalmology for sentiment and emotion analyses. We detailed the steps performed to quantify patients' and doctors' sentiments from web-based discussion forums: searching results, extracting a data corpus of threads and posts, cleaning the data, analyzing text using IBM Watson NLU, and aggregating and processing the important keywords from each post. Our goal was to explain these key steps and highlight the applicability of our methods to the field of medicine and the factors to consider in the process, notably the selection of search terms; understanding the latter's different linguistic usages (eg, idioms); the adequate consideration of different forums; and the establishment of robust criteria for data cleaning, aggregation, and grouping of posts and keywords (eg, lowercasing, punctuation removal, and lemmatization). Our approach highlights the importance of considering the unique structure of discussions within web-based health forums, distinguishing between physician and patient posts and analyzing idiomatic language usage to determine text relevance in infoveillance studies, which we found to be important steps not commonly detailed in previous studies of web-based health forums [39,40].

Medical Application

Analyses examining groupings (eg, administration; complications; procedures; medication and treatment; people; medical signs, symptoms, and diseases; time; and body parts), forum subgroups (eg, eye care, neurology, dermatology, thyroid disorders, multiple sclerosis, and undiagnosed symptoms), and patient-doctor interactions can enable researchers to provide key recommendations to physicians. In the oculoplastics data set, patients had a highly negative overall sentiment score and emotion score (anger, disgust, fear, and sadness) before the doctor replied (Table 2). To improve patient satisfaction, health care professionals can address their concerns by adapting their responses to the patients' sentiments and emotions. These sentiments and emotions can be further broken down by grouping and forums. Each grouping can be addressed with different solutions, such as reducing appointment and waiting time; explaining medical signs, symptoms, and diseases; and reassuring patients' concerns regarding specific procedures and body parts (Figure 4). Each forum's scores indicate how the corresponding health care team (eg, neurology, endocrinology, and ophthalmology) must communicate with patients to better manage different emotions, different emotions by predominantly addressing patients' sadness, disgust, fear, or even joy (Multimedia Appendix 4).

Challenges and Factors to Consider

Several issues must be carefully considered when gathering data from internet sources and unstructured free text to ensure relevance to the desired topic. First, the selection of the search terms is critical when analyzing web-based content. A deep understanding of the chosen field along with its related terms (eg, symptoms, complications, and subfields) is crucial to establish a complete list that encompasses all the possible relevant thread discussions. Second, a thorough understanding of the linguistic usages of the search terms is critical for establishing adequate data cleaning algorithms (eg, removal of threads containing the search terms exclusively used as idioms and consideration of human speech variance in the filtering algorithm). There are many eye-related idioms in the English language that must be considered when analyzing web-based text for ophthalmology-related insights (eg, "bat an eyelid"); every specialty will have its own unique set of idioms related to anatomical parts or functions (eg, "break my heart" and "take my breath away") that must be taken into consideration. The results can also differ according to the terms' specificity: broader terms (eg, eyelids, eyebrows, and oculoplastics) encompass the oculoplastics field, whereas more specific terms (eg, blepharitis, entropion, and ectropion) refer to specific medical conditions in this field. It is recommended to choose all relevant search terms (broad and specific) to ensure exhaustive results. However, a robust and tailored filtering algorithm must be established to ensure a relevant data set that is not influenced by the initial results returned by any proprietary search algorithm for any platform.

Indeed, every social media platform will have individual and proprietary search functions that may retrieve information irrelevant to the original query. Therefore, a careful and tailored process for further filtering is required to remove irrelevant results. Key decisions must be made on the filtering process (filtering by topic title, discussion thread, and/or individual post content). Establishing these filtering guidelines is crucial to ensure that the content of the posts selected is relevant and that the posts discarded do not contain relevant information. Basing the filtering algorithm on the relevancy of the thread topic allows for this methodology to be applied to many other social media platforms that often contain similar data structures (eg, on Facebook, Twitter, and Instagram, a main post (topic or title) is followed by comments (replies) related to the initial topic).

Furthermore, the scope of the search must also be evaluated. Depending on the topic selected, forums outside of those dedicated to the primary specialty may also need to be included. In our study, we considered a wide variety of MedHelp forums outside the eye care forums as oculoplastics is a field at the intersection of ophthalmology and plastic surgery. The *Eye Care* forum is only one of the 273 forums that contained our relevant threads and posts (ie, the *Cosmetic Surgery*, *Dermatology*, *Neurology*, and *Thyroid Disorders* forums). As we took all MedHelp forums into account during the extraction process, more constraints had to be established. For example, all forums related to animal care needed to be excluded.

After carefully selecting individual posts on which sentiment analysis is performed, the keywords extracted by the program will be numerous and lexically repetitive. Therefore, care must be taken to normalize the results originally sourced from free text. Using NLP tools to process and group the keywords with the same clinical meaning is a crucial step to ensure that the analysis is performed on uniform and clean data. To facilitate the grouping of related processed keywords, following a systematic method, such as ours (all keywords with a frequency greater than 500 and keyword categorization by 2 reviewers),

XSL•FO
RenderX

prevents biases from being induced into the sentiment analysis and results.

Limitations

Although the effects of users' spatiotemporal characteristics on sentiment analyses in MedHelp have not been evaluated yet, studies have shown that these features can bias the results of sentiment analysis derived from tweets. Gore et al showed that sentiment analysis can yield biased measures related to population demographics at the municipal, state, and national levels [41]. Another study demonstrated that an individual's location throughout the day can also affect their tweets' sentiment [42]. These issues can be addressed by assessing the population represented by posts on the web. In the case of Twitter, only 15% of adults on the web regularly use Twitter, and those aged 18-29 years and minorities tend to be more highly represented on Twitter than in the general population [43]. Although it is unclear what effect these spatial, temporal, and demographic effects may have on sentiment and emotion reflected in forum posts, they have the potential to affect these findings. We acknowledge that not all patients will rely on web-based forums to discuss their medical concerns or receive expert advice, especially the most vulnerable (older adults, minority, and socioeconomic groups).

Conclusions

Despite these limitations, the internet is a major source of health-related information that is underused [44]. In this paper, we describe an accessible, quick, and robust approach to sentiment analysis of patient data in social media that is relevant to a chosen medical topic, such as oculoplastics, and highlight the technical challenges encountered when preparing and analyzing the data. Regardless of the clinical questions examined, important factors to be considered during the application of this methodology include assessing the scope of the research; determining search terms and understanding their different linguistic usages; and implementing selection, filtering, and processing criteria for posts and keywords tailored to the results. This emerging methodology can be used as a valuable guide for clinicians and researchers who want to better understand patient attitudes toward and patient satisfaction with particular fields and procedures. The analysis of web-based forum discussions can be a quick, efficient, and robust method for gathering unstructured, diverse, and detailed opinions relevant to a chosen medical topic such as oculoplastics.

Acknowledgments

The authors thank Jerry Kurian for his expert opinion on the code and the reviewers for their valuable comments. Funding associated with this study includes T15 LM 007033 (SYW) and departmental support from the National Institute of Health-National Eye Institute Grant P30-EY026877 (SYW and AYW), as well as the unrestricted department grant from Research to Prevent Blindness, Inc (SYW and AYW). There are no commercial relationships to disclose.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial search terms.
[DOCX File , 31 KB - jmir_v23i5e20803_app1.docx]

Multimedia Appendix 2
List of patterns used to filter threads based on their titles or initial questions.
[DOCX File , 32 KB - jmir_v23i5e20803_app2.docx]

Multimedia Appendix 3
List of patterns used to filter posts based on their content.
[DOCX File , 31 KB - jmir_v23i5e20803_app3.docx]

Multimedia Appendix 4
Top 6 forums and their respective overall sentiment and emotion scores. The overall sentiment score reflects a positive, neutral, or negative sentiment, whereas the emotion score (anger, disgust, fear, joy, and sadness) represents how likely (%) the emotion is to be present. These forums had the highest number of posts and threads (displayed in the table).
[DOCX File , 16 KB - jmir_v23i5e20803_app4.docx]

References

1. Street RL, Haidet P. How well do doctors know their patients? Factors affecting physician understanding of patients' health beliefs. J Gen Intern Med 2011 Jan;26(1):21-27 [FREE Full text] [doi: 10.1007/s11606-010-1453-3] [Medline: 20652759]
2. Dawn AG, Lee PP. Patient expectations for medical and surgical care: a review of the literature and applications to ophthalmology. Surv Ophthalmol 2004;49(5):513-524. [doi: 10.1016/j.survophthal.2004.06.004] [Medline: 15325196]

XSL•FO
RenderX

3. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-04-24]
4. Global digital population as of January 2021. Statista. 2020. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2020-04-03]
5. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-04-24]
6. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-based content analysis of web-based health-related social media. J Med Internet Res 2016 Jun 13;18(6):e148 [FREE Full text] [doi: 10.2196/jmir.5327] [Medline: 27296242]
7. Pournaras E, Nikolic J, Omerzel A, Helbing D. Engineering democratization in internet of things data analytics. In: Proceedings of the IEEE 31st International Conference on Advanced Information Networking and Applications (AINA). 2017 Presented at: IEEE 31st International Conference on Advanced Information Networking and Applications (AINA); March 27-29, 2017; Taipei, Taiwan. [doi: 10.1109/aina.2017.15]
8. 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]
9. Das A, Faxvaag A. What influences patient participation in an online forum for weight loss surgery? A qualitative case study. Interact J Med Res 2014;3(1):e4 [FREE Full text] [doi: 10.2196/ijmr.2847] [Medline: 24509408]
10. Dosani S, Harding C, Wilson S. Online groups and patient forums. Curr Psychiatry Rep 2014 Nov;16(11):507 [FREE Full text] [doi: 10.1007/s11920-014-0507-3] [Medline: 25273668]
11. Haselmayer M, Jenny M. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. Qual Quant 2017;51(6):2623-2646 [FREE Full text] [doi: 10.1007/s11135-016-0412-4] [Medline: 29070915]
12. Ranco G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I. The effects of Twitter sentiment on stock price returns. PLoS One 2015;10(9):e0138441 [FREE Full text] [doi: 10.1371/journal.pone.0138441] [Medline: 26390434]
13. Htay SS, Lynn KT. Extracting product features and opinion words using pattern knowledge in customer reviews. ScientificWorldJournal 2013;2013:394758 [FREE Full text] [doi: 10.1155/2013/394758] [Medline: 24459430]
14. Garcia-Rudolph A, Laxe S, Saurí J, Bernabeu Guitart M. Stroke survivors on Twitter: sentiment and topic analysis from a gender perspective. J Med Internet Res 2019 Aug 26;21(8):e14077 [FREE Full text] [doi: 10.2196/14077] [Medline: 31452514]
15. Johnsen JK, Eggesvik TB, Rørvik TH, Hanssen MW, Wynn R, Kummervold PE. Differences in emotional and pain-related language in Tweets about dentists and medical doctors: text analysis of Twitter content. JMIR Public Health Surveill 2019 Feb 06;5(1):e10432 [FREE Full text] [doi: 10.2196/10432] [Medline: 30724738]
16. Talbot TB, Kalisch N, Christoffersen K, Lucas G, Forbell E. Stud Health Technol Inform 2016;220:407-413. [Medline: 27046614]
17. Denecke K, Deng Y. Sentiment analysis in medical settings: new opportunities and challenges. Artif Intell Med 2015 May;64(1):17-27. [doi: 10.1016/j.artmed.2015.03.006] [Medline: 25982909]
18. Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. JMIR Med Inform 2020 Jan 28;8(1):e16023 [FREE Full text] [doi: 10.2196/16023] [Medline: 32012057]
19. Wang SY, Hernandez-Boussard T, Chang RT, Pershing S. Understanding patient attitudes toward multifocal intraocular lenses in online medical forums through sentiment analysis. Stud Health Technol Inform 2019 Aug 21;264:1378-1382. [doi: 10.3233/SHTI190453] [Medline: 31438152]
20. 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]
21. Marshall SA, Yang CC, Ping Q, Zhao M, Avis NE, Ip EH. Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. Qual Life Res 2016 Mar;25(3):547-557 [FREE Full text] [doi: 10.1007/s11136-015-1156-7] [Medline: 26476836]
22. 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]
23. Williams SA, Dhillon S. Women's obstetric and reproductive health care discourse in online forums: perceived access and quality pre- and post-Affordable Care Act. Prev Med 2019 Jul;124:50-54. [doi: 10.1016/j.ypmed.2019.04.013] [Medline: 31028754]
24. Castaneda P, Sales A, Osborne NH, Corriere MA. Scope, themes, and medical accuracy of eHealth peripheral artery disease community forums. Ann Vasc Surg 2019 Jan;54:92-102. [doi: 10.1016/j.avsg.2018.09.004] [Medline: 30267913]
25. MedHelp. URL: https://www.medhelp.org/ [accessed 2021-05-06]
26. Hagan JC, Kutryb MJ. Internet eye questions. Ophthalmology 2009 Oct;116(10):2036. [doi: 10.1016/j.ophtha.2009.05.008] [Medline: 19800523]
27. Python. Python Software Foundation. URL: https://www.python.org/ [accessed 2021-05-05]
28. Richardson L. Beautiful Soup. 2007. URL: https://www.crummy.com/software/BeautifulSoup/bs4/doc/ [accessed 2020-08-13]
29. Hipp RD. SQLite. 2020. URL: https://www.sqlite.org/index.html [accessed 2021-04-24]

XSL•FO
RenderX

30. Trinh XV, Wang SY. Social Media Sentiment Emotion Analysis. Zenodo. URL: https://github.com/eyelovedata/social-media-sentiment-emotion-analysis [accessed 2020-08-25]
31. Bird S, Loper E, Klein E. Natural language processing with python. O'Reilly. 2009. URL: http://www.datascienceassn.org/sites/default/files/Natural%20Language%20Processing%20with%20Python.pdf [accessed 2020-08-13]
32. Watson Natural Language Understanding. IBM Corp. 2020. URL: https://www.ibm.com/ca-en/marketplace/natural-language-understanding [accessed 2020-04-03]
33. The Jupyter Notebook. The Jupyter Team. 2015. URL: https://jupyter-notebook.readthedocs.io/en/stable/ [accessed 2020-08-01]
34. Bird S, Loper E, Klein E. Natural language processing with python. Natural Language Toolkit. 2009. URL: https://www.nltk.org/ [accessed 2020-08-13]
35. NumPy. URL: https://numpy.org/ [accessed 2020-08-01]
36. Pandas. Qeios. URL: https://pandas.pydata.org/ [accessed 2021-04-24]
37. WordNet. Princeton University. 2010. URL: https://wordnet.princeton.edu/citing-wordnet [accessed 2021-04-24]
38. Oculoplastics keywords. URL: https://oculoplastics-keywords.herokuapp.com/ [accessed 2020-05-05]
39. Hua M, Sadah S, Hristidis V, Talbot P. Health effects associated with electronic cigarette use: automated mining of online forums. J Med Internet Res 2020 Jan 03;22(1):e15684 [FREE Full text] [doi: 10.2196/15684] [Medline: 31899452]
40. McRoy S, Rastegar-Mojarad M, Wang Y, Ruddy KJ, Haddad TC, Liu H. Assessing unmet information needs of breast cancer survivors: exploratory study of online health forums using text classification and retrieval. JMIR Cancer 2018 May 15;4(1):e10 [FREE Full text] [doi: 10.2196/cancer.9050] [Medline: 29764801]
41. 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]
42. 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]
43. Smith A, Brenner J. Twitter use 2012. 2012. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2021-04-24]
44. Carrillo-de-Albornoz J, Rodríguez Vidal J, Plaza L. Feature engineering for sentiment analysis in e-health forums. PLoS One 2018;13(11):e0207996 [FREE Full text] [doi: 10.1371/journal.pone.0207996] [Medline: 30496232]

Abbreviations

NLP: natural language processing
NLTK: Natural Language Toolkit
NLU: Natural Language Understanding

Edited by G Eysenbach; submitted 29.05.20; peer-reviewed by A Benis, R Gore; comments to author 12.08.20; revised version received 27.08.20; accepted 16.03.21; published 17.05.21.

Please cite as:
Nguyen AXL, Trinh XV, Wang SY, Wu AY
Determination of Patient Sentiment and Emotion in Ophthalmology: Infoveillance Tutorial on Web-Based Health Forum Discussions
J Med Internet Res 2021;23(5):e20803
URL: https://www.jmir.org/2021/5/e20803
doi:10.2196/20803
PMID:33999001

©Anne Xuan-Lan Nguyen, Xuan-Vi Trinh, Sophia Y Wang, Albert Y Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

Tutorial

Establishing and Facilitating Large-Scale Manuscript Collaborations via Social Media: Novel Method and Tools for Replication

Kimberly D Acquaviva[1], PhD, MSW, CSE, FNAP

School of Nursing, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Kimberly D Acquaviva, PhD, MSW, CSE, FNAP
School of Nursing
University of Virginia
4005 McLeod Hall
Charlottesville, VA, 22903
United States
Phone: 1 202 423 0984
Email: kda8xj@virginia.edu

Abstract

Background: Authorship teams in the health professions are typically composed of scholars who are acquainted with one another before a manuscript is written. Even if a scholar has identified a diverse group of collaborators outside their usual network, writing an article with a large number of co-authors poses significant logistical challenges.

Objective: This paper describes a novel method for establishing and facilitating large-scale manuscript collaborations via social media.

Methods: On September 11, 2020, I used the social media platform Twitter to invite people to collaborate on an article I had drafted. Anyone who wanted to collaborate was welcome, regardless of discipline, specialty, title, country of residence, or degree completion. During the 25 days that followed, I used Google Docs, Google Sheets, and Google Forms to manage all aspects of the collaboration.

Results: The collaboration resulted in the completion of 2 manuscripts in a 25-day period. The International Council of Medical Journal Editors authorship criteria were met by 40 collaborators for the first article ("Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions") and 35 collaborators for the second article ("The Benefits of Using Social Media as a Health Professional in Academia"). The authorship teams for both articles were notably diverse, with 17%-18% (7/40 and 6/35, respectively) of authors identifying as a person of color and/or underrepresented minority, 37%-38% (15/40 and 13/35, respectively) identifying as LGBTQ+ (lesbian, gay, bisexual, transgender, gender non-conforming, queer and/or questioning), 73%-74% (29/40 and 26/35, respectively) using she/her pronouns, and 20%-23% (9/40 and 7/35, respectively) identifying as a person with a disability.

Conclusions: Scholars in the health professions can use this paper in conjunction with the tools provided to replicate this process in carrying out their own large-scale manuscript collaborations.

(J Med Internet Res 2021;23(5):e25077) doi:10.2196/25077

KEYWORDS

social media; crowdsourcing; collaboration; health professions; medicine; scholarship; literature; research

Introduction

In the health professions, we typically collaborate with people we already know or are at least acquainted with. There is a practical reason for this: We have to know someone exists in order to think of them as a potential collaborator. However, this can lead to authorship teams that lack diversity. Even if a scholar has identified a diverse group of collaborators outside their usual network, writing an article with a large number of co-authors can be a logistical nightmare. This paper describes a novel method for establishing and facilitating large-scale manuscript collaborations via social media.

XSL•FO
RenderX

Methods

Finding Collaborators

On September 11, 2020, I posted a series of tweets (a "thread") on the social media platform Twitter, inviting people to collaborate with me on an article I had drafted [1]. Textbox 1 shows the transcript of the tweets in the thread.

The first tweet in the thread was shared with the >7400 people who follow me on Twitter; 2330 people engaged with the first tweet in the thread: 1991 people viewed the details about this tweet, 161 people clicked on my profile, 122 people clicked a heart icon to indicate they "liked" the tweet, 29 people replied to the tweet, and 27 people shared the tweet with their followers by retweeting it.

Less than 24 hours after I posted the thread inviting people to collaborate with me, 31 people had entered their names and affiliations onto the title page of the Google Doc to indicate their desire to serve as co-authors [2].

Textbox 1. Transcript of the tweets in the thread.

Publication opportunity for MD, NP, and RN tweeps! I'm writing a piece on how to document social media engagement as public scholarship on CVs and dossiers in medicine and nursing. Target Journal = @AcadMedJournal There's no widely-accepted format for how to do this so I thought it was about time we fixed that. :) Since the article is about social media engagement, I decided to take a risk and use social media to recruit co-authors and add to/revise the draft manuscript. Academic Medicine uses the ICMJE definition of authorship: "Authorship is based on (1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for "...important intellectual content, (3) final approval of the version to be published, and (4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately "...investigated and resolved. Authors must meet conditions 1, 2, 3, and 4." If you're willing to meet conditions 1, 2, 3, and 4, AND you're willing to turn around your edits in the next 7 days -- a timeframe I picked solely because I want to see how fast we can actually do this -- here's how you can co-author this paper with me. Step 1. Go to the Google Doc and type in your name and affiliation on the title page in a blue highlighted spots (you can add more spots if you need to). ** Rest assured your name will NOT be included on the submission unless you give approval of the final version.** Step 2. Make the article better. Work your magic. Don't insert comments -- add to or revise the actual text of the article. You have good ideas -- your ideas will make this article much better! Next Friday I'll take whatever edits have been made and will create a final version of the submission. I'll email it to everyone who listed their name on the title page. If you're happy with the final version and you meet the authorship criteria, bingo - you're a co-author. For a control-freak like me this is absolutely terrifying, BTW. So, here's the link to the Google Doc: [link to original Google Doc]. It's a rough draft, y'all -- like, ROUGH. If we pull this off, we can change the way CVs and dossiers look at med schools and nursing schools. The last person who did that was Ernest Boyer 30 years ago. How cool is that? In terms of the co-authors, I *really* want to have a diverse group. We can't have a bunch of white, cis, heterosexual, able-bodied folx writing this. If you're white, cis, heterosexual, and able-bodied, I'd still love for you to be a coauthor. Just saying that I'm hoping for a diverse group overall. I'm excited to see what comes of this. Thanks for being open to being part of this adventure! Here's the link to the Google Doc if you missed it earlier: [link to original Google Doc]. Thanks everybody!"

Coordinating the Collaboration

To manage the large number of potential co-authors, I created a Google Sheet and posted a link to it at the top of the title page of the Google Doc containing the draft article. With the link to the Google Sheet, I included a note encouraging co-authors to indicate in the Google Sheet whether they identified as someone who is LGBTQ+ (lesbian, gay, bisexual, transgender, gender non-conforming, queer and/or questioning), a person of color/under-represented minority, and/or a person with a disability because this would be helpful info for us to have collectively so that we have a sense of the diversity of our authorship team (I have created a blank version of the Co-Author Info Sheet for anyone interested in replicating the process described in this innovation report) [3]. There was no screening process for who could or could not join in the collaboration. Anyone who wanted to collaborate was welcome, regardless of discipline, specialty, title, country of residence, or degree completion.

Over the next 7 days, the article grew and changed considerably. Students, residents, fellows, and more senior scholars joined as collaborators, and no contributions to the article were viewed as more or less important based on the status of the collaborator making the contribution. Collaborators made edits to the article and inserted comments or questions that were then answered by other collaborators, occasionally in real time when multiple people were working on the document simultaneously. By September 18, 2020 — my original target date for completing the article — it was clear the article had expanded to have 2 foci rather than 1. I pulled content from Article #1 into a second Google Doc to build an outline for Article #2, then shared the link to the outline with all collaborators.

To keep things simple throughout the collaboration process, the first page of the original Google Doc was where I posted messages to the collaborators as well as links to the revised Article #1 (guidelines for documenting social media contributions) and what had become a draft of Article #2 (the benefits of social media for health professionals in academia). This approach preserved the comments made by collaborators on the original draft so that we could maintain a history of everyone's contributions.

Ensuring the Integrity of Authorship Designation

With such a large team of potential co-authors, I recognized that I would need to build several checkpoints into the collaboration process to ensure that individuals who met the International Council of Medical Journal Editors (ICMJE) criteria for authorship would be credited as such on the manuscript. The ICMJE defines authorship as follows: "The ICMJE recommends that authorship be based on the following 4 criteria: 1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND 2. Drafting the work or revising it critically for important intellectual content; AND 3. Final approval of the version to be published; AND 4. Agreement to be accountable for all aspects of the work in ensuring that

XSL•FO
RenderX

questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. In addition to being accountable for the parts of the work he or she has done, an author should be able to identify which co-authors are responsible for specific other parts of the work. In addition, authors should have confidence in the integrity of the contributions of their co-authors. All those designated as authors should meet all four criteria for authorship, and all who meet the four criteria should be identified as authors. Those who do not meet all four criteria should be acknowledged" [4].

To address this, I asked co-authors to describe their specific contributions to each manuscript in a Google Sheet; then, I sent them a "Co-Author Attestation Form" at 2 different points in the manuscript preparation process. This form asked them to attest to having met all 4 of the ICMJE authorship criteria. All members of the authorship team were able to see one another's contributions to the manuscript in the Google Doc as well as their descriptions of those contributions in the Google Sheet, so there was ample opportunity for concerns to be raised if an individual was not pulling their weight as a co-author. No such concerns were raised. In a few cases, individuals recognized during the co-author attestation checkpoints that they themselves did not meet the ICMJE authorship criteria, and as such, they requested being moved to the acknowledgements. Overall, this process seems to have been both efficient and effective.

Revising, Rewriting, and Negotiating

On September 18, 2020, one collaborator (J Mugele) took the lead in completely rewriting Article #1 to improve flow and cohesion. This collaborator's contributions were substantial in reshaping the way in which Article #1 was conceptualized, so I asked him to serve as second author. Although none of the co-authors expressed concerns about this, in hindsight, I should have asked the other collaborators if they supported this decision. At the time, I was struggling to manage what had grown into a team of 45 active, engaged collaborators; Mugele's leadership in revising the article was a life buoy, and I grabbed it. Mugele sent me the revised article on September 19, 2020, and I posted it to the rest of the collaborators later that day as a link from the original Google Doc. I asked all collaborators to make final edits by Friday, September 25, 2020 and asked them to indicate their preference in the Google Sheet for listing co-authors after the first 2 authors (Acquaviva and Mugele).

Between September 19, 2020 and September 25, 2020, I followed up with collaborators via Twitter direct message to remind them to finish entering their info on the Google Sheet. The vast majority of collaborators indicated a preference for listing co-authors after Acquaviva and Mugele in alphabetical order, so I edited the title page of the manuscript accordingly. On September 25, 2020, when the article was finalized and all edits had been received, I put together 2 different Google Forms to send to collaborators. I emailed "Co-Author Attestation v. 1" to collaborators who had already documented in the Google Sheet that they met the first 2 ICMJE authorship criteria. The form asked them to indicate they approved the final version of the article and met the last 2 ICMJE authorship criteria. I emailed "Co-Author Attestation v. 2" to collaborators who had not yet documented in the Google Sheet that they met the first 2 ICMJE authorship criteria. The form asked them to describe how they met the first criteria and then indicate whether they met the other 3 criteria. I have created blank versions of the Co-Author Attestation v.1 [5] and Co-Author Attestation v.2 [6] for readers who are interested in replicating the approach described in this innovation report. The last Co-Author Attestation form for Article #1 was received on September 29, 2020. In keeping with the practice I follow with all co-authored manuscripts I write, I ran the manuscript through iThenticate prior to submission to ensure that it was free of plagiarism. None was found.

Article #2

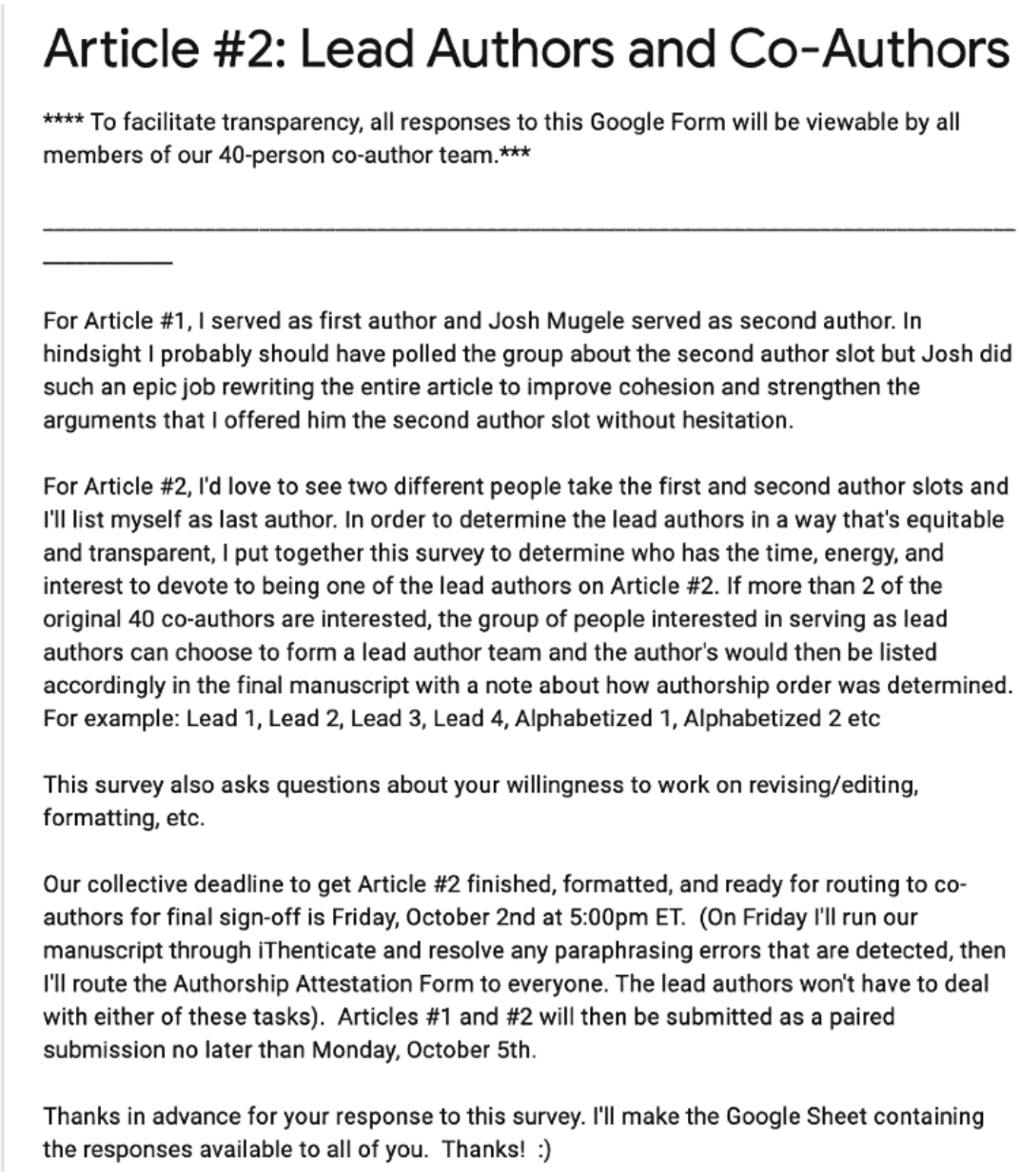
With Article #1 finalized, I then began the process of moving Article #2 toward completion. On September 27, 2020, I sent a Google Form to the 40 original co-authors and one individual who received an acknowledgement in Article #1 but made substantial contributions to the draft of Article #2. Because all 41 of these individuals had made significant contributions to the draft of Article #2, I wanted to give each person in the group the opportunity to take a lead role on the manuscript. The Google Form contained the message shown in Figure 1.

The Google Form contained 3 questions (Figure 2).

As responses to the Google Form were received, they automatically fed into a Google Sheet that was viewable by all members of the 41-person team. I created additional tabs in the Google Sheet labeled "Lead Author Team," "Revising/Editing Team," "Formatting Team," and "Move to Acknowledgements" and then copied individual responses into these sheets. On the evening of September 27, 2020, I posted the message shown in Figure 3 as both a comment and a cell in the Lead Author Team sheet.

XSL•FO
RenderX

Figure 1. Message contained within the authorship interest form.



Article #2: Lead Authors and Co-Authors

**** To facilitate transparency, all responses to this Google Form will be viewable by all members of our 40-person co-author team.***

For Article #1, I served as first author and Josh Mugele served as second author. In hindsight I probably should have polled the group about the second author slot but Josh did such an epic job rewriting the entire article to improve cohesion and strengthen the arguments that I offered him the second author slot without hesitation.

For Article #2, I'd love to see two different people take the first and second author slots and I'll list myself as last author. In order to determine the lead authors in a way that's equitable and transparent, I put together this survey to determine who has the time, energy, and interest to devote to being one of the lead authors on Article #2. If more than 2 of the original 40 co-authors are interested, the group of people interested in serving as lead authors can choose to form a lead author team and the author's would then be listed accordingly in the final manuscript with a note about how authorship order was determined. For example: Lead 1, Lead 2, Lead 3, Lead 4, Alphabetized 1, Alphabetized 2 etc

This survey also asks questions about your willingness to work on revising/editing, formatting, etc.

Our collective deadline to get Article #2 finished, formatted, and ready for routing to co-authors for final sign-off is Friday, October 2nd at 5:00pm ET. (On Friday I'll run our manuscript through iThenticate and resolve any paraphrasing errors that are detected, then I'll route the Authorship Attestation Form to everyone. The lead authors won't have to deal with either of these tasks). Articles #1 and #2 will then be submitted as a paired submission no later than Monday, October 5th.

Thanks in advance for your response to this survey. I'll make the Google Sheet containing the responses available to all of you. Thanks! :)

Over the next 7 days, the members of the Lead Author Team and Co-Author Team made considerable revisions to the manuscript. On October 3, 2020, I posted a link at the top of the manuscript Google Doc to a Google Form. The Google Form asked members of the authorship team if they happened to know the reference for 2 sentences that had been entered into the draft, as well as asked them about their contributions to the manuscript and their preferences for author order if they were a member of the Lead Author Team. Responses to the Google Form quickly yielded answers to the 2 reference questions but did little to answer the question of what order the lead author team members should be listed. On October 4, 2020, I asked a member of the Lead Author Team (Christopher Carroll) if he would take the lead on communicating with the rest of the lead author team via email to reach consensus on author order. Within 12 hours, the members of the Lead Author Team had reached consensus on the order their names should be listed on the final manuscript. Without exception, members of the Lead Author Team were generous in recognizing the contributions of other members and humble in conveying their own.

On October 5, 2020, I repeated the authorship attestation process followed with Article #1. I emailed the Co-Author Attestation Google Form along with a link to the final manuscript to all 39 collaborators. The last Co-Author Attestation form for Article #2 was received on October 5, 2020. In keeping with the practice I follow with all co-authored manuscripts I write, I ran the manuscript through iThenticate prior to submission to ensure that it was free of plagiarism. None was found.

XSL•FO
RenderX

Figure 2. Questions on the authorship interest form.

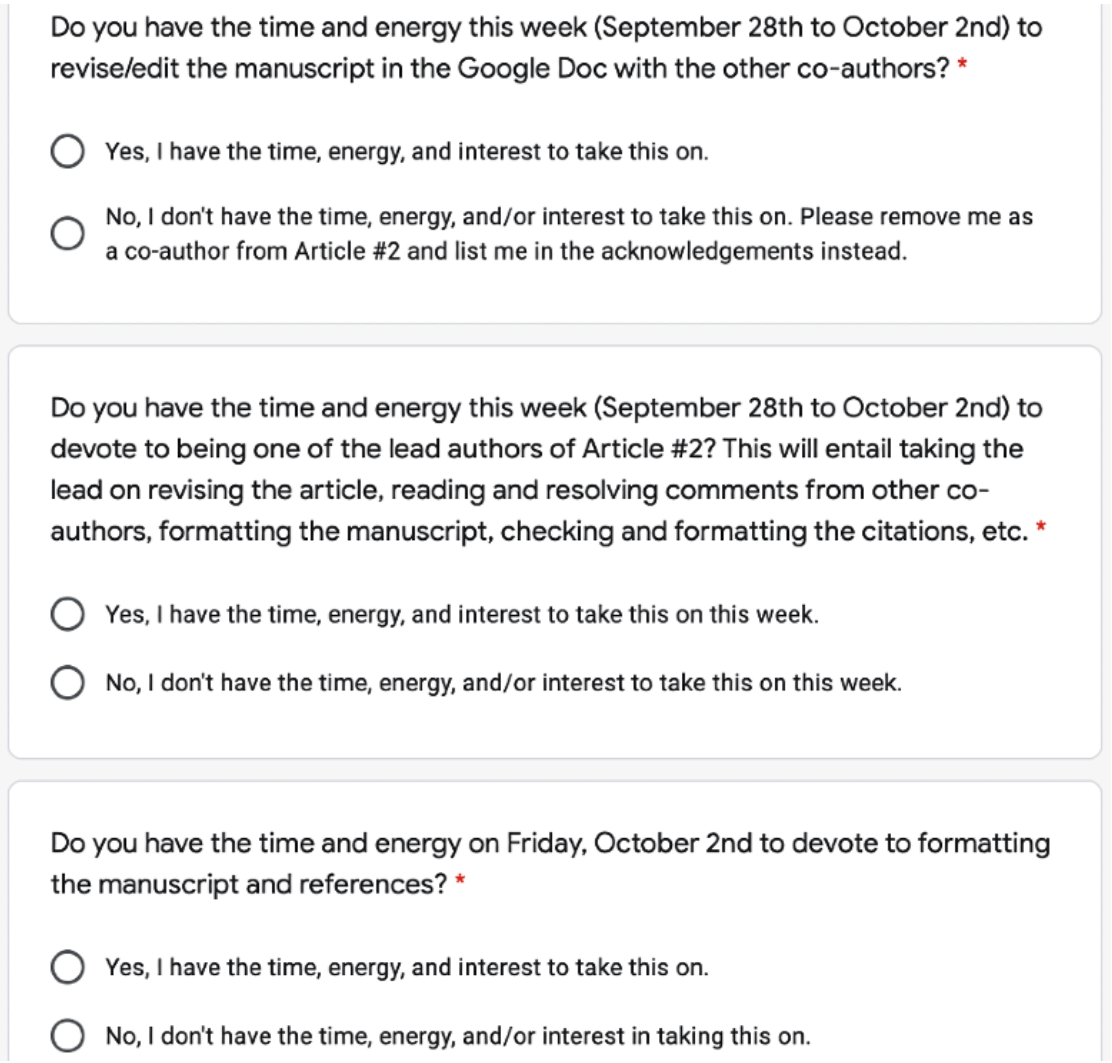
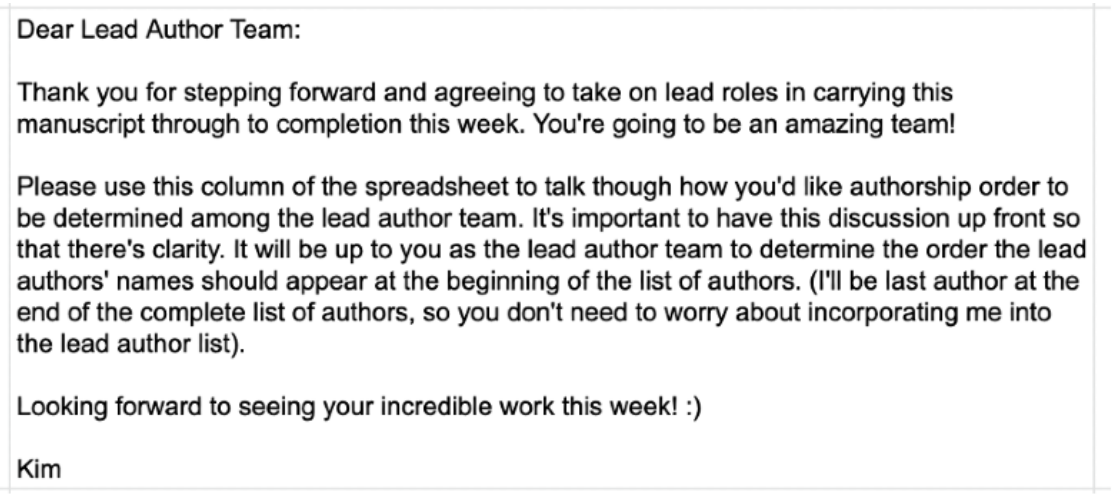


Figure 3. Message to the lead author team.



XSL•FO
RenderX

Results

Article #1: Documenting Social Media Engagement as Scholarship: New Model for Assessing Academic Accomplishment for the Health Professions [7]

Of the 45 collaborators who had originally entered their names into the title page of the Google Doc as potential co-authors, 40 collaborators ended up meeting the ICMJE authorship criteria, 4 collaborators decided that their contributions were more appropriate for recognition in an acknowledgement, and 1 collaborator dropped out because they were unable to contribute. The authorship team was notably diverse (Table 1).

Table 1. Authorship team composition for article #1.

Article #1	Identifies as a person of color and/or an under-represented minority, n (%)	Identifies as LGBTQ+[a], n (%)	Uses she/her pronouns	Identifies as a person with a disability
Authorship team (40 people)	7 (18)	15 (38)	29 (73)	9 (23)

[a]LGBTQ+: lesbian, gay, bisexual, transgender, gender non-conforming, queer and/or questioning.

To view the diverse array of degrees, licenses, and certifications held by the authorship team, you can view the article online [7].

After submitting the manuscript to Academic Medicine, we received a rejection within a matter of days. No feedback was provided. The authorship team revised the manuscript for submission to the Journal of Medical Internet Research (JMIR) and opted for open peer review. We received a revise-and-resubmit decision and then revised the manuscript to address the reviewers' comments and suggestions. The manuscript was then accepted for publication and published in December 2020. The day the article was published online, one of the members of the authorship team noticed that their name was not listed on the article: Apparently, I had made an error during the process of entering the metadata into the journal's online submission system. I was mortified. I immediately owned up to the error publicly on Twitter and contacted the authorship team to alert them directly. Because a correction to the list of authors in a journal article requires the approval of every member of the authorship team, I routed the corrigendum to every author for their review and signature. The error was corrected in the list of authors within a matter of days.

Article #2: The Benefits of Using Social Media as a Health Professional in Academia

Of the 41 collaborators from Article #1 who were surveyed at the beginning of the Article #2 revision process to determine their interest in serving as a lead author or co-author on the second manuscript, 2 individuals asked to be moved to the acknowledgements section of Article #2. Of the remaining 39 collaborators, 7 indicated interest in serving on the lead author team. When the manuscript was formatted and finalized and the Authorship Attestation Form was routed on October 3, 2020, 35 collaborators ended up meeting the ICMJE authorship criteria, and 4 collaborators agreed to be moved to the acknowledgements section. The authorship team for Article #2 was as diverse as Article #1, with marked diversity among the Lead Author Team in particular (Table 2).

Table 2. Authorship team composition for article #2.

Article #2	Identifies as a person of color and/or an under-represented minority, n (%)	Identifies as LGBTQ+[a], n (%)	Uses she/her pronouns, n (%)	Identifies as a person with a disability, n (%)
Lead authorship team (7 people)	2 (29)	3 (43)	3 (43)	2 (29)
Authorship team (35 people)	6 (17)	13 (37)	26 (74)	7 (20)

[a]LGBTQ+: lesbian, gay, bisexual, transgender, gender non-conforming, queer and/or questioning.

After submitting the manuscript to Academic Medicine, we received a rejection accompanied by helpful feedback for revising the paper for submission to another journal. The overarching theme running through the reviewers' comments was that we needed to focus the article more narrowly — a valid criticism with which the members of the authorship team agreed. The authorship team is in the process of revising the manuscript to address the issues raised by the Academic Medicine reviewers. Because Academic Medicine did not invite us to resubmit a revised manuscript, we plan to submit the paper to another journal that is better aligned with the subject matter.

Relationship Between the Three Manuscripts

Congruent with the ICMJE's "Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals" [4], this manuscript is not "…reporting work that has already been reported in large part in a published article or is contained in or closely related to another paper that has been submitted or accepted for publication elsewhere." This manuscript is narrowly focused on describing the method used to establish and facilitate a large-scale collaboration using social media. The manuscript titled "Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions" [7] presents the guidelines created through the process described in this manuscript, while the manuscript titled "The Benefits of Using Social Media as a Health Professional in Academia" (unpublished) examines the benefits of social media engagement.

Conclusion

This paper and the 2 manuscripts described therein are evidence that it is possible to establish and facilitate large-scale manuscript collaborations via social media. Open collaboration

XSL•FO
RenderX

on manuscripts in the health professions is one way to ensure diverse authorship teams and facilitate collaboration across disciplines. Scholars can use this innovation report in conjunction with the tools provided to replicate this process in carrying out their own large-scale manuscript collaborations.

Acknowledgments

The author wishes to thank the collaborators who participated in the process of developing the 2 manuscripts described in this article: Natasha Abadilla, BA; Jocelyn C Anderson, PhD, RN; Tyler Adamson, MPH; Samantha L Bernstein, MSN, RNC-OB, IBCLC; Rakhee K Bhayani, MD; Annina E Büchi, MD, MSc; Darcy Burbage, DNP, RN, AOCN, CBCN; Christopher L Carroll, MD; Julie SW Childers, MD, PhD; Samantha P Davis, MS, RRT, RRT-NPS, CHSE; Natasha Dhawan, MD; Alice Eaton, BSc; Kim English, RN, BScN, MN; Jennifer T Grier, PhD; Mary K Gurney, RPh, PhD, BCPA; Emily S Hahn, MD; Heather Haq, MD, MHS; Brendan Huang, BS; Shikha Jain, MD, FACP; Jin Jun, PhD, RN; S Alexander Kemery, PhD, RN; Wesley T Kerr, MD, PhD; Timothy Keyes; Amelia R Kirby, BA; Rebecca S Koszalinski, PhD, RN, CRRN, CMSRN; Marion Leary, RN, MSN, MPH; Chrystal L Lewis, PhD, RN; Mollie Marr, BFA; Ajay Major, MD, MBA; Jason V Meisel, PMHNP-BC; Josh Mugele, MD; Erika A Petersen, MD, FAANS, FACS; Barak Raguan, MD; Allison Rhodes, BA; Deborah D Rupert, MS, MA; Nadia A Sam-Agudu, MD, CTropMed; Naledi Saul, MPM; Jarna Shah, MD; Lisa Kennedy Sheldon, PhD, ANP-BC, AOCNP, FAAN; Christian T Sinclair, MD; Kerry Spencer, PhD; Natalie H Strand, MD; Carl G Streed Jr, MD, MPH; and Avery M Trudell, BSA.

Conflicts of Interest

None declared.

References

1. Acquaviva K. Thread by @kimacquaviva. Thread Reader App. 2020 Sep 11. URL: https://perma.cc/3FES-JF79 [accessed 2021-05-08]
2. Acquaviva K. Tweet: It's only been 21 hrs since. Twitter. 2020 Sep 12. URL: https://perma.cc/G8LW-NCLT [accessed 2021-05-08]
3. Acquaviva K. Co-Author Info Sheet. Google Sheets. 2020 Oct 16. URL: https://perma.cc/BQN6-E94R [accessed 2021-05-08]
4. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals. International Committee of Medical Journal Editors (ICMJE). 2019 Dec. URL: http://www.icmje.org/icmje-recommendations.pdf [accessed 2021-05-08]
5. Acquaviva K. Co-Author Attestation v1. Google Forms. 2020 Sep 26. URL: https://perma.cc/YD32-TW85 [accessed 2021-05-08]
6. Acquaviva K. Co-Author Attestation v2. Google Forms. 2020 Sep 26. URL: https://perma.cc/X9ZC-PWDX [accessed 2021-05-08]
7. Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, et al. Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions. J Med Internet Res 2020 Dec 2;22(12):e25070 [FREE Full text] [doi: 10.2196/25070] [Medline: 33263554]

Abbreviations

ICMJE: International Council of Medical Journal Editors
LGBTQ+: lesbian, gay, bisexual, transgender, gender non-conforming, queer and/or questioning

Edited by R Kukafka; submitted 16.10.20; peer-reviewed by R Halkes; comments to author 20.11.20; revised version received 24.11.20; accepted 30.04.21; published 17.05.21.

Please cite as:
Acquaviva KD
Establishing and Facilitating Large-Scale Manuscript Collaborations via Social Media: Novel Method and Tools for Replication
J Med Internet Res 2021;23(5):e25077
URL: https://www.jmir.org/2021/5/e25077
doi:10.2196/25077
PMID:33999002

©Kimberly D Acquaviva. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(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

Initial Outcomes of Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy Tailored to Public Safety Personnel: Longitudinal Observational Study

Heather D Hadjistavropoulos[1], PhD; Hugh C McCall[1], MA; David L Thiessen[2], BSc; Ziyin Huang[3], HBSc; R Nicholas Carleton[1,3], PhD; Blake F Dear[4], PhD; Nickolai Titov[4], PhD

[1]Department of Psychology, University of Regina, Regina, SK, Canada

[2]Department of Mathematics and Statistics, University of Regina, Regina, SK, Canada

[3]PSPNET, University of Regina, Regina, SK, Canada

[4]eCentreClinic, Department of Psychology, Macquarie University, Sydney, Australia

Corresponding Author:
Heather D Hadjistavropoulos, PhD
Department of Psychology
University of Regina
Administration-Humanities Building, AH 345
3737 Wascana Parkway
Regina, SK, S4S 0A2
Canada
Phone: 1 306 585 5133
Email: Heather.Hadjistavropoulos@uregina.ca

Abstract

Background: Canadian public safety personnel (PSP) experience high rates of mental health disorders and face many barriers to treatment. Internet-delivered cognitive behavioral therapy (ICBT) overcomes many such barriers, and is effective for treating depression, anxiety, and posttraumatic stress disorder (PTSD) symptoms.

Objective: This study was designed to fill a gap in the literature regarding the use of ICBT tailored specifically for PSP. We examined the effectiveness of a tailored ICBT program for treating depression, anxiety, and PTSD symptoms among PSP in the province of Saskatchewan.

Methods: We employed a longitudinal single-group open-trial design (N=83) with outcome measures administered at screening and at 8 weeks posttreatment. Data were collected between December 5, 2019 and September 11, 2020. Primary outcomes included changes in depression, anxiety, and PTSD symptoms. Secondary outcomes included changes in functional impairment; symptoms of panic, social anxiety, and anger; as well as treatment satisfaction, working alliance, and program usage patterns.

Results: Clients reported large symptom reductions on measures of depression and anxiety, as well as moderate reductions on measures of PTSD and secondary symptoms, except for social anxiety. Most clients who reported symptoms above clinical cut-offs on measures of depression, anxiety, and PTSD during screening experienced clinically significant symptom reductions. Results suggested good engagement, treatment satisfaction, and working alliance.

Conclusions: Tailored, transdiagnostic ICBT demonstrated promising outcomes as a treatment for depression, anxiety, and PTSD among Saskatchewan PSP and warrants further investigation.

Trial Registration: Clinicaltrials.gov NCT04127032; https://www.clinicaltrials.gov/ct2/show/NCT04127032

(J Med Internet Res 2021;23(5):e27610) doi:10.2196/27610

KEYWORDS

internet; cognitive behavior therapy; anxiety; depression; posttraumatic stress disorder; public safety personnel; CBT; internet-based cognitive behavioral therapy; ICBT; PTSD; outcome; diagnosis; longitudinal; observational; literature; effectiveness

XSL•FO
RenderX

Introduction

Public safety personnel (PSP) include border security personnel, communications officials (eg, emergency response dispatchers), correctional workers, firefighters, paramedics, police officers, and others whose work keeps communities safe [1]. Through their work, PSP frequently experience or witness potentially traumatic events such as motor vehicle collisions and acts of violence [2-4]. Research shows that such events put individuals at risk of developing several emotional disorders, including posttraumatic stress disorder (PTSD), major depressive disorder, and panic disorder [5-10]. In Canada, a large survey study indicated that 44.5% of PSP report clinically significant symptoms of one or more mental health disorders [11]. PSP have also reported high rates of mental health problems in other countries [12-14]. PSP face many barriers to treatment, such as stigma, distance from services, and time constraints [15]. Stigma appears to be a particularly prohibitive barrier to treatment among PSP [15,16].

Internet-delivered cognitive behavioral therapy (ICBT) is a treatment wherein traditional cognitive behavioral therapy skills are taught via online learning platforms and is poised to help overcome these barriers to treatment for PSP. ICBT can be accessed privately and conveniently, and it is cost-effective because it requires less therapist time per client than traditional face-to-face treatments [17,18]. Meta-analyses have shown that ICBT is effective for treating symptoms of a range of mental disorders, including depression, anxiety, and PTSD [19,20]. There have been several recent calls for internet interventions to be tailored to the unique needs of their target users to facilitate greater treatment engagement [21,22]. As one group stated, this process begins with establishing "a deep understanding of user needs and preferences, and actively involving users in design processes from the outset" [21]. Tailoring ICBT may be particularly important for PSP, who have reported several unique needs such as the need for flexible treatment timelines, diverse preferences for frequency of therapist support, specialized therapists with knowledge of unique PSP issues and culture, and a focus on exposures to potentially psychologically traumatic events [15]. Canadian PSP appear receptive to therapist-supported ICBT, which was ranked as their most preferred mental health treatment after face-to-face psychological services [23].

The Government of Canada recently announced a $10 million investment into developing, delivering, and evaluating ICBT tailored specifically for PSP [24]. Our research unit, PSPNET, was contracted to carry out this work. We conducted extensive interviews with PSP stakeholders to better understand their needs and preferences [15] and, drawing upon their feedback, tailored a transdiagnostic ICBT program that previously demonstrated positive outcomes for a range of conditions to meet the needs of PSP. The tailored course of treatment is called the PSP Wellbeing Course, which is described in detail below.

This paper presents the initial outcomes of the PSP Wellbeing Course among the first 83 PSP enrolled in the course during the first 7 months of availability in Saskatchewan. The study objectives were to evaluate the initial (a) effectiveness, (b) treatment satisfaction and working alliance, and (c) program usage patterns of the PSP Wellbeing Course. We hypothesized high completion rates, high satisfaction, and moderate to large effect sizes for symptom change. This study is the first, as far as we are aware, to evaluate the tailored ICBT approach for treating mental health problems among PSP.

This study's sample represents only the first third of the sample described in our trial protocol (N=250; Clinicaltrials.gov NCT04127032). This early evaluation of outcomes is important for three reasons. First, it is important to ensure that the development and delivery of the PSP Wellbeing Course is a fruitful use of Canadian tax dollars. Second, evaluation of early outcomes is consistent with the learning health system approach, which is characterized by a continuous cycle of gathering data from practice, transforming data into knowledge, and implementing knowledge into practice [25]. Third, our preliminary findings may be helpful for others involved in the development, provision, evaluation, and funding of mental health services for PSP.

Methods

Study Design

We employed a longitudinal single-group open-trial design. This study was approved by the local institutional research ethics board (2019-157) and registered on Clinicaltrials.gov (NCT04127032). This article presents the initial study outcomes for the first 83 clients who enrolled in the PSP Wellbeing Course (one third of the total expected sample of 250). We have followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement in preparing the article [26].

Setting

This study was performed in the Canadian province of Saskatchewan. There was no specialized ICBT for PSP in Saskatchewan prior to this study. ICBT has been available to Saskatchewan residents through the Online Therapy Unit [27] since 2010, but the numbers of PSP using this service have not been systematically tracked.

Clients

All clients who enrolled in the PSP Wellbeing Course between December 5, 2019 and July 2, 2020 consented to our use of their data for research and were included in this study. Enrollment in the PSP Wellbeing Course required clients to (a) be 18 years or older, (b) currently or formerly employed as a PSP (as defined above), (c) have access to a computer and internet service, (d) provide an emergency medical contact, and (e) reside in the province of Saskatchewan (although we later began offering the PSP Wellbeing Course in English and French to residents of Quebec as well). Prospective clients were excluded and referred to appropriate services if they (a) reported a high suicide risk; (b) had attempted suicide or had been hospitalized for a high suicide risk within the past year; (c) reported a primary problem with psychosis, mania, or alcohol and drugs; (d) reported receiving psychological treatment more than twice per month; or (e) reported concerns about ICBT and requested referral to other services in the community.

XSL•FO
RenderX

Intervention

The PSP Wellbeing Course was adapted from a transdiagnostic ICBT program called the Wellbeing Course, which was developed for use in the general population by eCentre Clinic at Macquarie University, Australia [28-32]. The PSP Wellbeing Course is designed to treat symptoms of depression, anxiety, and PTSD. The course consists of five sequential lessons, which are presented in a slideshow format and focus on (1) introducing the cognitive behavioral model and symptom identification, (2) monitoring and challenging maladaptive thoughts, (3) managing physical symptoms, (4) graded exposure, and (5) relapse prevention. Lesson content is relayed through instructive text, diagrams, and illustrative case stories about PSP. Clients are also given access to downloadable materials, weekly homework assignments, and supplementary resources related to issues not addressed in the lessons (eg, assertiveness, communication, sleep problems). Therapists hold graduate degrees in psychology or social work and provide support to clients by phone or secure email up to twice per week, depending on each client's preference. Each week, clients complete symptom questionnaires and a reflection questionnaire including both open-ended and closed-ended questions designed to gauge response to the treatment program (eg, what clients worked on; any challenges they encountered; and ratings of helpfulness, understanding, and difficulty). The course is designed to be completed in 8 weeks, but clients are provided with therapist support for up to 16 weeks based on symptoms and preference, and are given access to the course materials for 1 year. We tailored the PSP Wellbeing Course using extensive feedback from PSP stakeholders gleaned through 112 interviews [15] and 132 survey responses [23] (eg, replacing generic case stories and examples used to illustrate treatment concepts with case stories and examples about PSP).

Procedure

Prospective clients completed a two-stage eligibility screening process consisting of an online screening questionnaire (demographics, clinical history, and the measures described below) and a screening interview by phone. Clients then began the PSP Wellbeing Course. Clients were automatically administered brief weekly questionnaires and a longer battery of questionnaires at 8 weeks, as detailed below. Questionnaires were embedded in the course. Members of our research team followed up with clients by phone or email to encourage them to complete posttreatment measures.

Measures

Primary Outcome Measures

Patient Health Questionnaire-9

The Patient Health Questionnaire-9 (PHQ-9) measures symptoms of depression and includes one item to assess suicidality [33]. It includes nine items rated from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 27. A score greater than 9 is typically used to identify a likely diagnosis of major depressive disorder [34]. The PHQ-9 has demonstrated excellent psychometric properties [33,35,36]. Cronbach α for the PHQ-9 in this study ranged from .84 to .88.

Generalized Anxiety Disorder-7

Generalized Anxiety Disorder-7 (GAD-7) is a brief measure of general anxiety symptoms [37]. It includes seven items rated from 0 (not at all) to 3 (nearly every day) for a total score ranging from 0 to 21. The GAD-7 has demonstrated strong psychometric properties, and a score of 10 or greater is typically used to identify a likely diagnosis of generalized anxiety disorder [36,37]. Cronbach α for the GAD-7 in this study ranged from .86 to .92.

Posttraumatic Stress Disorder Checklist for DSM-5

The Posttraumatic Stress Disorder Checklist for the DSM-5 (PCL-5) measures symptoms of PTSD and includes 20 items rated from 0 (not at all) to 4 (extremely) [38]. A cut-off of 33 was recommended by several groups to indicate a likely diagnosis of PTSD [39,40]. The PCL-5 has shown good psychometric properties [39]. Cronbach α for the PCL-5 in this study ranged from .95 to .96.

Secondary Outcome Measures

Panic Disorder Severity Scale Self-Report

The Panic Disorder Severity Scale Self-Report (PDSS-SR) is a psychometrically sound 7-item measure assessing symptoms of panic disorder [41]. Items are rated from 0 to 4 for a total score of 0 to 28. It is common practice to employ a cut-off score of 8 on the PDSS-SR [42-44], based on the cut-off initially established for the nonself-report PDSS [45]. Cronbach α for the PDSS-SR in this study was .89-.91.

Six-Item Social Phobia Scale and Six-Item Social Interaction Anxiety Scale

The 6-item Social Phobia Scale (SPS-6) and 6-item Social Interaction Anxiety Scale (SIAS-6) are brief measures of social anxiety that are psychometrically sound when administered together or separately [46]. Items are rated from 0 to 4 for total scores of 0 to 24 on each measure or a combined total of 0 to 48, with a clinical cut-off of 2 on the SPS-6 and of 7 on the SIAS-6 [46]. Cronbach α for the measures combined in the current study was .91 and .94.

Dimensions of Anger Reactions Scale

The Dimensions of Anger Reactions Scale (DAR-5) is a 5-item measure of anger that has demonstrated good psychometric properties [47]. Scores range from 5 to 25 with a cut-off of 12 [48]. Cronbach α for the DAR-5 in this study was .88 and .89.

Sheehan Disability Scale

The Sheehan Disability Scale (SDS) is a psychometrically sound 3-item measure assessing respondents' functioning in their home lives, social lives, and work/school lives [49].

Treatment Satisfaction Questionnaire

We administered a bespoke questionnaire to explore treatment satisfaction.

Working Alliance Inventory-Short Revised

The Working Alliance Inventory-Short Revised (WAI-SR) is a 12-item measure of working alliance with goal, task, and bond subscales that has demonstrated good psychometric properties [50,51]. The Cronbach α for the WAI-SR in this study was .94

XSL•FO
RenderX

for the goal subscale, .93 for the task subscale, and .95 for the bond subscale.

Administration of Measures

At screening and posttreatment, clients were administered the PHQ-9, GAD-7, PLC-5, PDSS-SR, SPS-6, SIAS-6, DAR-5, and SDS. Additionally, at posttreatment, clients were administered the Treatment Satisfaction Questionnaire and WAI-SR. Certain measures were administered weekly during treatment, and at 6 months and 1 year postenrollment. These time points are not included in the present analyses but will be reported in the future once more data are available.

Analyses

Following intention-to-treat principles, we first created multiple imputations for missing data using predictive mean matching. Imputations for missing data were modeled controlling for lesson completion rates, pretreatment symptom measures, and, if available, weekly symptom measures. We used generalized estimating equations (GEEs) to model the change in symptoms from pretreatment to posttreatment [52]. A gamma distribution with a log-link function was used to model changes between pretreatment and posttreatment as proportional and to accommodate the skewed response distribution [53]. We ran the imputations and GEE models in R version 4.0.3 using the mice [54] and geepack [55] packages.

The effectiveness of the PSP Wellbeing Course was evaluated by examining Hedges *g* and proportional reductions in symptoms. Hedges *g* is calculated as the difference in means between pretreatment and posttreatment divided by the pooled SD and multiplied by a small-sample bias correction. It is interpreted similarly to Cohen *d* but was considered to be a more appropriate measure of effect size for this study given the small size of our clinical subsamples. Some researchers have recommended calculating Hedges *g* using the SD of the control or pretreatment group instead of the pooled SD when pretreatment and posttreatment SDs are unequal [56]. Indeed, we noted that in the subsets of clients scoring above clinical cut-offs at pretreatment, the pretreatment SDs were much lower than the posttreatment SDs. However, because these subsets represented a restricted range of pretreatment scores, a greater SD at posttreatment would be expected—even in the absence of a treatment effect—due to regression toward the mean [57]. Therefore, we used pooled SDs in calculating Hedges *g* in the clinical subsets to avoid overestimating effect sizes.

We also examined clinically significant deterioration and recovery for the PHQ-9, GAD-7, and PCL-5. Following the UK National Health Service guidelines, deterioration was defined as an increase of at least 6 points on the PHQ-9 [58]. Following precedents set in previous research using the same outcome measures, deterioration was defined as an increase of 5 points on the GAD-7 [59,60] or 10 points on the PCL-5 [61]. Recovery was defined as a decrease of at least 6 points on the PHQ-9, 5 points on the GAD-7, or 10 points on the PCL-5 that resulted in the patient moving from the clinical range to the nonclinical cut-off scores described in the Measures section above. We examined the uptake, treatment satisfaction and acceptability, and program usage patterns using descriptive statistics in SPSS version 23.

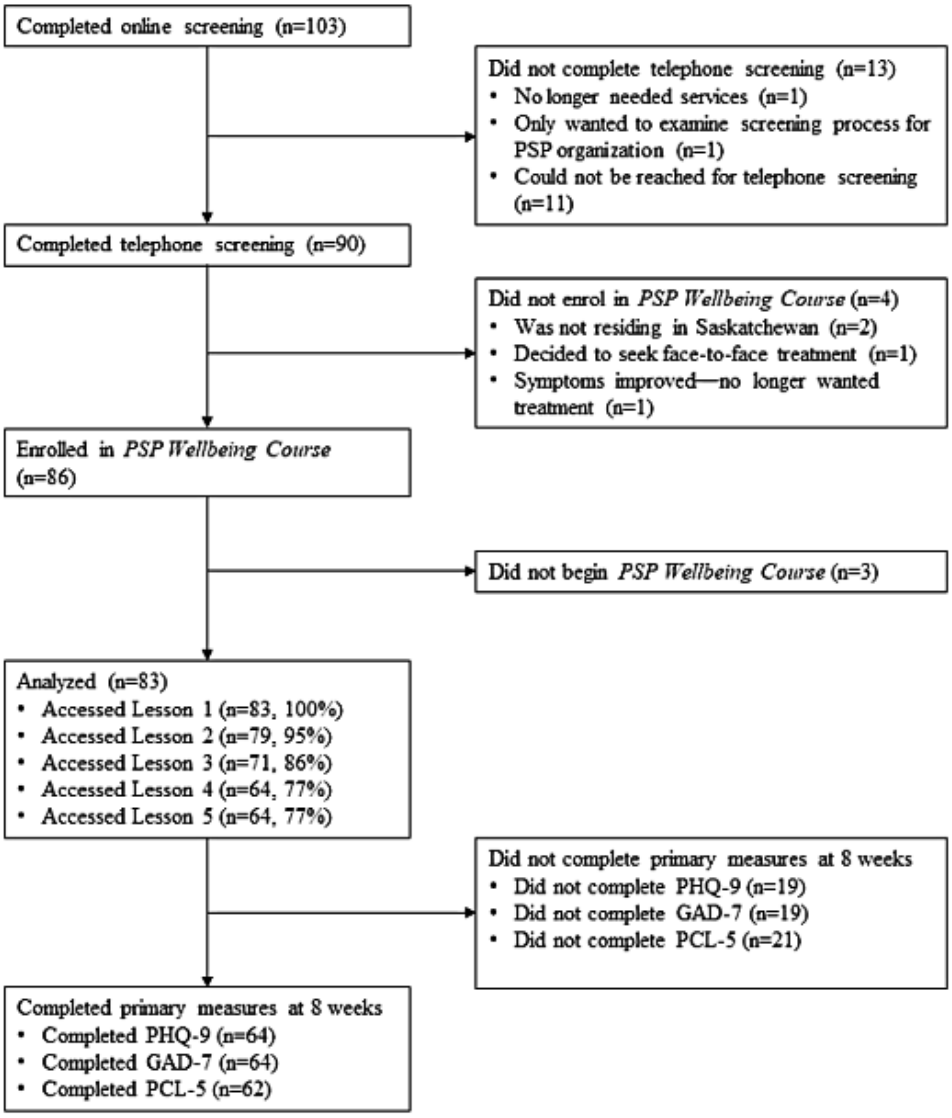
Results

Client Characteristics

There were 83 clients included in this study. The flow of clients through the study process is shown in Figure 1. Client characteristics are summarized in Table 1.

XSL•FO
RenderX

Figure 1. Flowchart displaying client enrollment, program usage, and completion of posttreatment measures. GAD-7: Generalized Anxiety Disorder-7; PCL-5: Posttraumatic Stress Disorder Checklist for the DSM-5; PHQ-9: Patient Health Questionnaire-9; PSP: public safety personnel.



XSL•FO
RenderX

Table 1. Client demographics (N=83).

Characteristic	Value
Age (years), mean (SD)	40.25 (10.95)
Gender, n (%)	
Woman	45 (54)
Man	37 (45)
Nonbinary	1 (1)
Ethnicity, n (%)	
White	73 (88)
First Nations, Inuit, or Metis	7 (8)
Other	3 (4)
Married or common law, n (%)	
Yes	53 (64)
No	30 (36)
Children, n (%)	
Yes	44 (53)
No	39 (47)
Community size, n (%)	
Population of 100,000 or greater	43 (52)
Population under 100,000	40 (48)
PSP[a] sector, n (%)	
Police	24 (29)
Corrections	9 (11)
Emergency medical service	29 (35)
Fire	6 (7)
Dispatch/communications	5 (6)
Other	10 (12)

[a]PSP: public safety personnel.

Effectiveness

The effect sizes for each scale, along with the percentage changes and reliable change rates for primary symptom measures are detailed in Table 2. The results indicated large effect sizes on the PHQ-9 and GAD-7, and a moderate effect size on the PCL-5. When examining effect sizes for each primary measure only among clients reporting clinically significant symptoms on that measure at screening, we found large effect sizes on all three measures. In general, we found moderate effect sizes on secondary symptom measures in the entire sample and moderate to large effect sizes when examining only clients reporting clinically significant symptoms at screening; however, the results indicated small effect sizes on the combined SIAS-6 and SPS-6 among all clients and for those with clinically significant SIAS-6 and SPS-6 scores at screening.

XSL•FO
RenderX

Table 2. Effectiveness of the PSP Wellbeing Course.

Questionnaire	Pretreatment, mean (SD)	Posttreatment, mean (SD)	Hedges *g* effect size (95% CI) from pretreatment	Percentage change (95% CI) in estimated mean from pretreatment	Reliable deterioration (%)	Reliable recovery (%)
Total sample (N=83)						
PHQ-9[a]	11.09 (5.59)	6.81 (5.33)	0.78 (0.46 to 1.09)	39 (26 to 49)	4.4	29.7
GAD-7[b]	10.32 (5.44)	6.01 (4.82)	0.83 (0.52 to 1.15)	42 (30 to 52)	1.7	37.6
PCL-5[c]	29.82 (19.18)	20.39 (17.60)	0.51 (0.20 to 0.82)	32 (15 to 45)	6.8	25.1
PDSS-SR[d]	5.20 (5.09)	3.17 (4.52)	0.42 (0.11 to 0.73)	39 (15 to 56)	N/A[e]	N/A
SIAS-6/SPS-6[f]	9.76 (9.22)	9.19 (9.58)	0.06 (–0.24 to 0.36)	6 (–20 to 26)	N/A	N/A
DAR-5[g]	10.31 (4.39)	8.59 (3.79)	0.42 (0.11 to 0.72)	17 (8 to 25)	N/A	N/A
SDS[h] Family	4.98 (2.73)	3.69 (2.84)	0.46 (0.15 to 0.77)	26 (10 to 39)	N/A	N/A
SDS Social	5.03 (2.94)	3.82 (2.92)	0.41 (0.10 to 0.72)	24 (9 to 37)	N/A	N/A
SDS Work	4.54 (3.02)	3.40 (3.05)	0.37 (0.07 to 0.68)	25 (7 to 40)	N/A	N/A
SDS Total	14.54 (7.81)	10.80 (7.76)	0.48 (0.17 to 0.79)	26 (11 to 38)	N/A	N/A
Clinical sample[i]						
PHQ-9 (≥10, n=48)	14.90 (3.82)	8.91 (5.46)	1.25 (0.82 to 1.70)	40 (28 to 51)	2.9	51.3
GAD-7 (≥10, n=46)	14.22 (3.24)	8.08 (5.11)	1.41 (0.97 to 1.87)	43 (31 to 53)	2.5	56.8
PCL-5 (≥33, n=35)	48.63 (10.66)	32.96 (17.63)	1.05 (0.56 to 1.56)	32 (17 to 45)	7.8	59.5
PDSS-SR (≥8, n=23)	12.14 (2.94)	7.00 (6.18)	1.03 (0.42 to 1.66)	42 (15 to 61)	N/A	N/A
SIAS-6/SPS-6 (SIAS≥7 SPS≥2, n=27)	20.52 (7.54)	18.48 (10.23)	0.22 (–0.31 to 0.76)	10 (–13 to 28)	N/A	N/A
DAR-5 (≥12, n=25)	15.84 (3.21)	11.76 (4.80)	0.97 (0.39 to 1.57)	26 (12 to 37)	N/A	N/A
SDS Family (≥3, n=66)	6.03 (1.93)	4.23 (2.84)	0.73 (0.38 to 1.09)	30 (15 to 42)	N/A	N/A
SDS Social (≥3, n=62)	6.34 (2.11)	4.45 (2.88)	0.74 (0.38 to 1.11)	30 (16 to 42)	N/A	N/A
SDS Work (≥3, n=58)	5.98 (2.40)	4.23 (3.00)	0.64 (0.27 to 1.01)	29 (13 to 43)	N/A	N/A
SDS Total (≥9, n=62)	17.95 (5.74)	12.35 (7.75)	0.81 (0.45 to 1.18)	31 (17 to 43)	N/A	N/A

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder Questionnaire-7.

[c]PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5.

[d]PDSS-SR: Panic Disorder Severity Scale Self-Report.

[e]N/A: not applicable; reliable recovery and deterioration were only calculated for primary measures because the magnitude of the score change required to determine reliable score changes for several secondary measures was unknown.

[f]SIAS-6/SPS-6: 6-item Social Interaction Anxiety Scale/6-item Social Phobia Scale.

[g]DAR-5: Dimensions of Anger Reactions-5.

[h]SDS: Sheehan Disability Scale.

[i]For each measure, estimated means, effects sizes, and percentage change are calculated based on those who scored in the clinical range on the measure at pretreatment.

Treatment Satisfaction and Working Alliance

Of the 62 clients who completed the WAI-SR and questions related to treatment satisfaction, 54/62 (86%) indicated that the PSP Wellbeing Course had increased or greatly increased their confidence in their ability to manage their symptoms, and 61/62 (98%) indicated that the PSP Wellbeing Course was worth their time. Mean scores on the bond, goal, and task subscales of the WAI-SR were 16.11 (SD 4.64), 14.50 (SD 4.72), and 14.63 (SD 4.07), respectively. To our knowledge, there are no formal guidelines for interpreting WAI-SR scores, but these participant scores fall in the “high/positive” range according to the interpretation of one group of ICBT researchers [62].

Program Usage Patterns

All 83 clients included in the analyses accessed Lesson 1 of the PSP Wellbeing Course. Most clients also accessed Lesson 2 (79/83, 95%), Lesson 3 (71/83, 86%), Lesson 4 (64/83, 77%),

XSL•FO
RenderX

and Lesson 5 (64/83, 77%). Of the 60 clients who completed 8-week measures, at the time they did so, all had accessed Lesson 1 and Lesson 2, and most had accessed Lesson 3 (58/60, 97%), Lesson 4 (54/60, 90%), and Lesson 5 (45/60, 75%). Across the entire sample, most clients who accessed each lesson also accessed the accompanying DIY guides (92% or more for each lesson), FAQs (84% or more), and stories about PSP (84% or more). Many clients accessed additional resources; the most frequently accessed resources were on relationships with significant others (46/83, 55%), anger (42/83, 51%), communication skills (40/83, 48%), and problem solving and worry time (40/83, 48%). Most clients opted to receive therapist support once per week (74/83, 89%), but some opted to receive therapist support twice per week (6/83, 7%) or only on an as-needed basis (3/83, 4%). On average, clients sent 4.98 messages (SD 5.53) to their therapists and received 9.80 (SD 4.71) messages.

Some clients reported using the PSP Wellbeing Course concurrently with other treatments. At prescreen, some clients reported currently taking one or more medications for their mental health (18/83, 22%) or receiving mental health care from one or more providers (19/83, 23%), the most common of which were family doctors (13/83, 16%), psychologists (6/83, 7%), psychiatrists (4/83, 5%), and counsellors (3/83, 4%). Per the eligibility criteria described above, none of these clients were receiving mental health treatments from these providers more than twice per month. Several clients (4/83, 5%) indicated that they were on a waitlist to see a mental health care provider.

Discussion

Principal Findings

Canadian PSP experience high rates of mental health problems and face several barriers to mental health care [15,16]. ICBT is a promising solution for effectively managing symptoms of common mental disorders [19,20] in a private and convenient fashion that allows clients to overcome key barriers for care [15,17]. PSPNET has developed a tailored ICBT program for PSP called the PSP Wellbeing Course. This article presents the initial program outcomes.

The results indicated that the PSP Wellbeing Course is effective for treating symptoms of depression, anxiety, PTSD, panic disorder, and anger, and also improves clients' functioning across three domains of life. The course appeared to be only slightly effective for treating symptoms of social anxiety, but this should be interpreted with caution because the subsample reporting clinically significant social anxiety was small (n=27). Nonetheless, in keeping with the learning health system approach [25], PSPNET will explore the possibility of providing additional resources to help clients manage social anxiety symptoms. This is important because approximately 15% of Canadian PSP—and 33% in this study's treatment-seeking sample—struggle with clinically significant social anxiety [11]. Results concerning treatment satisfaction, working alliance, and engagement were promising. Although in our interviews with PSP stakeholders [15], most had expressed a preference for a flexible treatment duration, it was encouraging that most clients completed all lessons within 8 weeks. It was also very encouraging that working alliance was high because many PSP we interviewed previously [15] reported negative attitudes toward mental health care providers (eg, distrust, discomfort, beliefs that service providers do not understand PSP). Most clients accessed multiple additional resources, suggesting that they found them helpful and lending some support to recommendations for modularized, optional treatment elements in internet interventions [21].

Descriptively, these outcomes are comparable to the most recent published outcomes of the Wellbeing Course offered by the Online Therapy Unit to the general population of Saskatchewan [63]. Effect sizes, percentage changes, and reliable change rates were comparable for all available symptom measures with the possible exception of the SIAS-6/SPS-6, which appeared to show better outcomes for the Wellbeing Course. However, mean pretreatment social anxiety symptoms were considerably less severe in the present sample than in the Online Therapy Unit's clients [63], and the smaller change in symptoms found in this study may be due, in part, to a floor effect. Of note, the present findings show that the PSP Wellbeing Course was moderately effective for treating anger, which has not been measured in research on the Wellbeing Course. PSP participating in this study reported concurrent use of psychotropic medication less frequently than clients of the Wellbeing Course [63]. Results concerning treatment satisfaction, working alliance, and engagement were similar across the two courses.

Limitations

This study has several limitations. First, the data were collected during the COVID-19 pandemic, and the impact of the pandemic on outcomes is unclear. Second, there was no control condition and therefore no direct evidence that symptom change can be attributed to the PSP Wellbeing Course. However, it has been argued that the literature on digital mental health interventions would benefit from more research evaluating such interventions in real-world settings [64,65], as we did in this study. Third, at 8 weeks, we were missing PHQ-9 and GAD-7 data for 23% of clients and PCL-5 data for 25% of clients. A conservative approach was used to manage missing data, but it is possible that clients for whom data were missing may represent a unique subset of clients (eg, clients who experienced less positive outcomes).

Future Directions

Further research can extend these findings in several ways. This article describes the results for only the first third of the sample (83 of 250) specified in our trial protocol; full results will be reported when available, including outcomes at 1-year follow-up and for PSPNET clients in Quebec. Further research will also be required to explore demographic and clinical predictors of treatment response for the PSP Wellbeing Course and to identify new ways to tailor and improve ICBT for PSP. Finally, although this study suggests that the PSP Wellbeing Course benefits the PSP who access it, economic evaluation will be required in the future.

Conclusions

Preliminary results indicate that the PSP Wellbeing Course is a promising and effective method for treating symptoms of

XSL•FO
RenderX

depression, anxiety, and PTSD among Saskatchewan PSP. Our results also suggest that the course improved clients' functioning and reduced their symptoms of panic and anger, but not social anxiety. Clients reported good working alliance and treatment satisfaction, and they demonstrated good engagement. Consistent with the concept of learning health systems, PSPNET will continue applying research results to iteratively improve tailored ICBT services, such as by investigating ways to better treat symptoms of social anxiety among PSP. Our findings may also be helpful for other groups involved in researching, providing, funding, or otherwise supporting mental health care for this unique and underserved population.

Acknowledgments

We would like to acknowledge the contributions of all PSP who have participated and the members of the PSPNET team for their support with this project. We would also like to acknowledge the many stakeholders and organizations who have assisted us in tailoring the PSP Wellbeing Course and distributing information about PSPNET to PSP. Acknowledgements are due to the Online Therapy Unit for their support in setting up PSPNET. The authors are indebted to Information Services at the University of Regina, especially Max Ivanov, for his pivotal role in developing the Online Therapy Unit platform that is used to host PSPNET. This research was performed by PSPNET, which is funded by the Canadian Government's Ministry of Public Safety and Emergency Preparedness. The funders had no involvement in the design of the study, or in the collection, analysis, and interpretation of the data. NT and BD are funded by the Australian Government to operate the MindSpot Clinic. NT and BF developed the Wellbeing Course but derive no financial benefit from it.

Conflicts of Interest

HH is the Director of PSPNET and Executive Director of the Online Therapy Unit, and declares funding as noted above. BFD and NT are authors and developers of the Wellbeing Course, but do not derive any personal or financial benefit from it. The other authors have no conflicts of interest to declare.

References

1. Oliphant R. Healthy minds, safe communities: supporting our public safety officers through a national strategy for operational stress injuries. Canadian House of Commons. 2016. URL: https://www.ourcommons.ca/Content/Committee/421/SECU/Reports/RP8457704/securp05/securp05-e.pdf [accessed 2021-04-28]
2. Carleton RN, Afifi TO, Taillieu T, Turner S, Krakauer R, Anderson GS, et al. Exposures to potentially traumatic events among public safety personnel in Canada. Can J of Behav Sci 2019 Jan;51(1):37-52. [doi: 10.1037/cbs0000115]
3. Galatzer-Levy IR, Madan A, Neylan TC, Henn-Haase C, Marmar CR. Peritraumatic and trait dissociation differentiate police officers with resilient versus symptomatic trajectories of posttraumatic stress symptoms. J Trauma Stress 2011 Oct 06;24(5):557-565 [FREE Full text] [doi: 10.1002/jts.20684] [Medline: 21898602]
4. Komarovskaya I, Maguen S, McCaslin SE, Metzler TJ, Madan A, Brown AD, et al. The impact of killing and injuring others on mental health symptoms among police officers. J Psychiatr Res 2011 Oct;45(10):1332-1336 [FREE Full text] [doi: 10.1016/j.jpsychires.2011.05.004] [Medline: 21658717]
5. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Washington DC: American Psychiatric Association; 2013.
6. Kilpatrick DG, Resnick HS, Milanak ME, Miller MW, Keyes KM, Friedman MJ. National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. J Trauma Stress 2013 Oct;26(5):537-547 [FREE Full text] [doi: 10.1002/jts.21848] [Medline: 24151000]
7. Nixon RD, Resick PA, Griffin MG. Panic following trauma: the etiology of acute posttraumatic arousal. J Anxiety Disord 2004 Jan;18(2):193-210. [doi: 10.1016/s0887-6185(02)00290-6]
8. O'Donnell ML, Creamer M, Pattison P. Posttraumatic stress disorder and depression following trauma: understanding comorbidity. Am J Psychiatry 2004 Aug;161(8):1390-1396. [doi: 10.1176/appi.ajp.161.8.1390] [Medline: 15285964]
9. Perkonigg A, Kessler R, Storz S, Wittchen HU. Traumatic events and post-traumatic stress disorder in the community: prevalence, risk factors and comorbidity. Acta Psychiatr Scand 2000 Jan;101(1):46-59. [doi: 10.1034/j.1600-0447.2000.101001046.x] [Medline: 10674950]
10. Shalev AY, Freedman S, Peri T, Brandes D, Sahar T, Orr SP, et al. Prospective study of posttraumatic stress disorder and depression following trauma. Am J Psychiatry 1998 May;155(5):630-637. [doi: 10.1176/ajp.155.5.630] [Medline: 9585714]
11. Carleton RN, Afifi TO, Turner S, Taillieu T, Duranceau S, LeBouthillier DM, et al. Mental disorder symptoms among public safety personnel in Canada. Can J Psychiatry 2018 Jan 28;63(1):54-64 [FREE Full text] [doi: 10.1177/0706743717723825] [Medline: 28845686]
12. Courtney JA, Francis AJP, Paxton SJ. Caring for the country: fatigue, sleep and mental health in Australian rural paramedic shiftworkers. J Community Health 2013 Mar 29;38(1):178-186. [doi: 10.1007/s10900-012-9599-z] [Medline: 22843315]

XSL•FO
RenderX

13. Maia DB, Marmar CR, Metzler T, Nóbrega A, Berger W, Mendlowicz MV, et al. Post-traumatic stress symptoms in an elite unit of Brazilian police officers: prevalence and impact on psychosocial functioning and on physical and mental health. J Affect Disord 2007 Jan;97(1-3):241-245. [doi: 10.1016/j.jad.2006.06.004] [Medline: 16859752]
14. Stanley IH, Boffa JW, Hom MA, Kimbrel NA, Joiner TE. Differences in psychiatric symptoms and barriers to mental health care between volunteer and career firefighters. Psychiatry Res 2017 Jan;247:236-242. [doi: 10.1016/j.psychres.2016.11.037] [Medline: 27930964]
15. McCall HC, Beahm JD, Fournier AK, Burnett JL, Carleton RN, Hadjistavropoulos HD. Stakeholder perspectives on internet-delivered cognitive behavioural therapy for public safety personnel: A qualitative analysis. Can J Behav Sci 2020 Aug 27:online ahead of print. [doi: 10.1037/cbs0000242]
16. Ricciardelli R, Carleton RN, Mooney T, Cramm H. "Playing the system": Structural factors potentiating mental health stigma, challenging awareness, and creating barriers to care for Canadian public safety personnel. Health (London) 2020 May 16;24(3):259-278. [doi: 10.1177/1363459318800167] [Medline: 32283964]
17. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
18. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
19. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul 16;64(7):465-470. [doi: 10.1177/0706743719839381] [Medline: 31096757]
20. Sijbrandij M, Kunovski I, Cuijpers P. Effectiveness of internet-delivered cognitive behavioral therapy for posttraumatic stress disorder: a systematic review and meta-analysis. Depress Anxiety 2016 Sep;33(9):783-791. [doi: 10.1002/da.22533] [Medline: 27322710]
21. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016 Apr 18;7:65. [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
22. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
23. McCall HC, Sison AP, Burnett JL, Beahm JD, Hadjistavropoulos HD. Exploring perceptions of internet-delivered cognitive behaviour therapy among public safety personnel: informing dissemination efforts. Int J Environ Res Public Health 2020 Aug 19;17(17):6026 [FREE Full text] [doi: 10.3390/ijerph17176026] [Medline: 32824994]
24. Supporting Canada's public safety personnel: an action plan on post-traumatic stress injuries. Public Safety Canada. 2019. URL: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2019-ctn-pln-ptsi/index-en.aspx [accessed 2021-04-27]
25. Menear M, Blanchette M, Demers-Payette O, Roy D. A framework for value-creating learning health systems. Health Res Policy Syst 2019 Aug 09;17(1):79 [FREE Full text] [doi: 10.1186/s12961-019-0477-3] [Medline: 31399114]
26. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
27. Online Therapy Unit. URL: https://www.onlinetherapyuser.ca/ [accessed 2020-04-28]
28. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: A randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
29. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot Clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
30. Fogliati V, Dear B, Staples L, Terides M, Sheehan J, Johnston L, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided internet-delivered treatment for panic disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2016 Apr;39:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2016.03.005] [Medline: 27003376]
31. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
32. Dear B, Staples L, Terides M, Fogliati V, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for Social Anxiety Disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]

XSL•FO
RenderX

33. 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]
34. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Mar 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
35. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008 Jan;58(546):32-36 [FREE Full text] [doi: 10.3399/bjgp08X263794] [Medline: 18186994]
36. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
37. Spitzer RL, Kroenke K, Williams JBW, 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]
38. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
39. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
40. Wortmann JH, Jordan AH, Weathers FW, Resick PA, Dondanville KA, Hall-Clark B, et al. Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. Psychol Assess 2016 Nov;28(11):1392-1403. [doi: 10.1037/pas0000260] [Medline: 26751087]
41. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety 2002 Jun 05;15(4):183-185. [doi: 10.1002/da.10049] [Medline: 12112724]
42. Christensen H, Batterham PJ, Grant JB, Griffiths KM, Mackinnon AJ. A population study comparing screening performance of prototypes for depression and anxiety with standard scales. BMC Med Res Methodol 2011 Nov 22;11(1):154 [FREE Full text] [doi: 10.1186/1471-2288-11-154] [Medline: 22103584]
43. Ruwaard J, Broeksteeg J, Schrieken B, Emmelkamp P, Lange A. Web-based therapist-assisted cognitive behavioral treatment of panic symptoms: a randomized controlled trial with a three-year follow-up. J Anxiety Disord 2010 May;24(4):387-396. [doi: 10.1016/j.janxdis.2010.01.010] [Medline: 20227241]
44. Allen AR, Newby JM, Mackenzie A, Smith J, Boulton M, Loughnan SA, et al. Internet cognitive-behavioural treatment for panic disorder: randomised controlled trial and evidence of effectiveness in primary care. BJPsych Open 2016 Mar;2(2):154-162 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001826] [Medline: 27703768]
45. Shear M, Rucci P, Williams J, Frank E, Grochocinski V, Vander Bilt J, et al. Reliability and validity of the Panic Disorder Severity Scale: replication and extension. J Psychiatr Res 2001 Sep;35(5):293-296. [doi: 10.1016/s0022-3956(01)00028-0]
46. Peters L, Sunderland M, Andrews G, Rapee RM, Mattick RP. Development of a short form Social Interaction Anxiety (SIAS) and Social Phobia Scale (SPS) using nonparametric item response theory: the SIAS-6 and the SPS-6. Psychol Assess 2012 Mar;24(1):66-76. [doi: 10.1037/a0024544] [Medline: 21744971]
47. Hawthorne G, Mouthaan J, Forbes D, Novaco RW. Response categories and anger measurement: do fewer categories result in poorer measurement?: development of the DAR5. Soc Psychiatry Psychiatr Epidemiol 2006 Mar;41(2):164-172. [doi: 10.1007/s00127-005-0986-y] [Medline: 16362166]
48. Forbes D, Alkemade N, Mitchell D, Elhai JD, McHugh T, Bates G, et al. Utility of the Dimensions of Anger Reactions-5 (DAR-5) scale as a brief anger measure. Depress Anxiety 2014 Mar 25;31(2):166-173. [doi: 10.1002/da.22148] [Medline: 23801571]
49. Sheehan D. The Sheehan Disability Scales. In: The anxiety disease and how to overcome it. New York: Charles Scribner and Sons; 1983.
50. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychother Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
51. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
52. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
53. Karin E, Dear BF, Heller GZ, Gandy M, Titov N. Measurement of symptom change following web-based psychotherapy: statistical characteristics and analytical methods for measuring and interpreting change. JMIR Ment Health 2018 Jul 12;5(3):e10200 [FREE Full text] [doi: 10.2196/10200] [Medline: 30001999]
54. Buuren SV, Groothuis-Oudshoorn K. Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]

XSL•FO
RenderX

55. Halekoh U, Højsgaard S, Yan J. The R Package geepack for Generalized Estimating Equations. J Stat Soft 2006;15(2):1-11 [FREE Full text] [doi: 10.18637/jss.v015.i02]
56. Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 2016 Nov 30;5(10):3-8. [doi: 10.3102/0013189x005010003]
57. Nesselroade JR, Stigler SM, Baltes PB. Regression toward the mean and the study of change. Psychol Bull 1980;88(3):622-637. [doi: 10.1037/0033-2909.88.3.622]
58. The improving access to psychological therapies manual: Appendices and helpful resources. National Collaborating Centre for Mental Health. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2018/06/iapt-manual-appendices-and-helpful-resources-v3.pdf [accessed 2021-04-28]
59. 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]
60. Titov N, Dear BF, Nielssen O, Wootton B, Kayrouz R, Karin E, et al. User characteristics and outcomes from a national digital mental health service: an observational study of registrants of the Australian MindSpot Clinic. Lancet Digit Health 2020 Nov;2(11):e582-e593 [FREE Full text] [doi: 10.1016/S2589-7500(20)30224-7] [Medline: 33103097]
61. Byllesby BM, Dickstein BD, Chard KM. The probability of change versus dropout in veterans receiving cognitive processing therapy for posttraumatic stress disorder. Behav Res Ther 2019 Dec;123:103483. [doi: 10.1016/j.brat.2019.103483] [Medline: 31710956]
62. Jasper K, Weise C, Conrad I, Andersson G, Hiller W, Kleinstäuber M. The working alliance in a randomized controlled trial comparing Internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. Internet Interv 2014 Apr;1(2):49-57. [doi: 10.1016/j.invent.2014.04.002]
63. Hadjistavropoulos H, Peynenburg V, Thiessen D, Nugent M, Adlam K, Owens K, et al. A pragmatic factorial randomized controlled trial of transdiagnostic internet-delivered cognitive behavioural therapy: Exploring benefits of homework reflection questionnaires and twice-weekly therapist support. Internet Interv 2020 Dec;22:100357 [FREE Full text] [doi: 10.1016/j.invent.2020.100357] [Medline: 33335846]
64. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
65. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 01;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]

Abbreviations

DAR-5: Dimensions of Anger Reactions Scale
GAD-7: Generalized Anxiety Disorder-7
GEE: generalized estimating equations
ICBT: internet-delivered cognitive behavioral therapy
PCL-5: Posttraumatic Stress Disorder Checklist for the DSM-5
PDSS-SR: Panic Disorder Severity Scale Self-Report
PHQ-9: Patient Health Questionnaire-9
PSP: public safety personnel
PTSD: posttraumatic stress disorder
SDS: Sheehan Disability Scale
SIAS-6: 6-item Social Interaction Anxiety Scale
SPS-6: 6-item Social Phobia Scale
WAI-SR: Working Alliance Inventory-Short Revised

Edited by G Eysenbach; submitted 30.01.21; peer-reviewed by K Mathiasen; comments to author 25.02.21; revised version received 01.03.21; accepted 11.04.21; published 05.05.21.

Please cite as:
Hadjistavropoulos HD, McCall HC, Thiessen DL, Huang Z, Carleton RN, Dear BF, Titov N
Initial Outcomes of Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy Tailored to Public Safety Personnel: Longitudinal Observational Study
J Med Internet Res 2021;23(5):e27610
URL: https://www.jmir.org/2021/5/e27610
doi:10.2196/27610
PMID:33949959

XSL•FO
RenderX

©Heather D Hadjistavropoulos, Hugh C McCall, David L Thiessen, Ziyin Huang, R Nicholas Carleton, Blake F Dear, Nickolai Titov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2021. 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 Internet-Based Cognitive Behavioral Therapy for Depression: Implementation Cost-Effectiveness Study

Jordi Piera-Jiménez[1,2,3], BSc, MSc; Anne Etzelmueller[4], BSc, MSc; Spyros Kolovos[5], PhD; Frans Folkvord[1,6], PhD; Francisco Lupiáñez-Villanueva[1,7], PhD

[1]Open Evidence Research Group, Universitat Oberta de Catalunya, Barcelona, Spain

[2]Servei Català de la Salut, Barcelona, Spain

[3]Digitalization for the Sustainability of the Healthcare System, Sistema de Salut de Catalunya, Barcelona, Spain

[4]GET.ON Institute GmbH/HelloBetter, Hamburg, Germany

[5]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[6]Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

[7]Department of Information and Communication Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Jordi Piera-Jiménez, BSc, MSc
Open Evidence Research Group
Universitat Oberta de Catalunya
Rambla del Poblenou 156
Barcelona, 08018
Spain
Phone: 34 651041515
Email: jpieraj@uoc.edu

Abstract

Background: Major depressive disorder is a chronic condition; its prevalence is expected to grow with the aging trend of high-income countries. Internet-based cognitive-behavioral therapy has proven efficacy in treating major depressive disorder.

Objective: The objective of this study was to assess the cost-effectiveness of implementing a community internet-based cognitive behavioral therapy intervention (Super@, the Spanish program for the MasterMind project) for treating major depressive disorder.

Methods: The cost-effectiveness of the Super@ program was assessed with the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing tool, using a 3-state Markov model. Data from the cost and effectiveness of the intervention were prospectively collected from the implementation of the program by a health care provider in Badalona, Spain; the corresponding data for usual care were gathered from the literature. The health states, transition probabilities, and utilities were computed using Patient Health Questionnaire–9 scores.

Results: The analysis was performed using data from 229 participants using the Super@ program. Results showed that the intervention was more costly than usual care; the discounted (3%) and nondiscounted incremental cost-effectiveness ratios were €29,367 and €26,484 per quality-adjusted life-year, respectively (approximately US $35,299 and $31,833, respectively). The intervention was cost-effective based on the €30,000 willingness-to-pay threshold typically applied in Spain (equivalent to approximately $36,060). According to the deterministic sensitivity analyses, the potential reduction of costs associated with intervention scale-up would reduce the incremental cost-effectiveness ratio of the intervention, although it remained more costly than usual care. A discount in the incremental effects up to 5% exceeded the willingness-to-pay threshold of €30,000.

Conclusions: The Super@ program, an internet-based cognitive behavioral therapy intervention for treating major depressive disorder, cost more than treatment as usual. Nevertheless, its implementation in Spain would be cost-effective from health care and societal perspectives, given the willingness-to-pay threshold of €30,000 compared with treatment as usual.

(J Med Internet Res 2021;23(5):e27410) doi:10.2196/27410

KEYWORDS

digital health; telemedicine; eHealth; e–mental health; internet-based cognitive behavioral therapy; depression; iCBT; implementation; cost-effectiveness; cognitive behavioral therapy; CBT; cost

XSL•FO
RenderX

Introduction

Population aging is a global trend and is expected to be one of the most significant societal challenges worldwide in upcoming years [1]. The profound impact that this aging trend is likely to cause on our societies and economies has prompted significant efforts in turning the challenges of this scenario into opportunities for rethinking the way we design and organize our society, including the delivery of health and social care services [2-5].

Depression is a significant contributor to morbidity during entire lifespans and has been among the 3 leading nonfatal causes of disability globally for nearly three decades [6]. Although often underdiagnosed, depression is the most prevalent mental health condition among adult population and across cultural settings resulting in an aggregate point prevalence of 12.9%, 1-year prevalence of 7.2%, and lifetime prevalence of 10.8% (years 1994-2014) [7-9].

The burden of depression is specifically high among the elderly, irrespective of the presence of cognitive impairment, particularly in long-term care settings [8,10,11]. Various factors may increase the risk of depression among older people, including physiological factors (eg, cardiovascular disease, diabetes, or immunological changes) and psychosocial factors (eg, low economic status, social isolation, or relocation) [12-14]. Once established, depression in older people increases the risk of suicide and may trigger dementia [10].

While the efficacy of psychotherapy in the treatment of depression has been proven [15], the availability of evidence-based interventions constitutes a persistent challenge given the lack and unequal distribution of qualified practitioners, delayed provision of treatment, and inadequacy of treatment [16,17]. Given the limitations and health care costs associated with treating depression (eg, US $7638, according to a study conducted in Singapore [18]), there is growing interest in alternative therapies to routine care. Among them, a plethora of internet-based cognitive behavioral therapies for depression treatment have been introduced, many showing efficacy in treating major depressive disorder [19-21]. Although costs associated with the implementation of these therapies have been assessed, most studies are based on descriptive approaches, and formal cost-effectiveness analysis of internet-based cognitive behavioral therapy interventions are scarce [22].

While randomized controlled trials and accompanying cost-effectiveness analysis can be considered the gold standard in exploring the cost-effectiveness of mental health interventions, the idealized and controlled nature of these trials limits the generalizability of findings to routine care populations [23-25]. Establishing the cost-effectiveness of an intervention and its implementation under routine care conditions is an important part of the evaluation before wide-scale adoption. So far, establishing the cost-effectiveness of implementation projects in routine care provides a methodological challenge.

MasterMind was a European cofunded project aimed at scaling-up the implementation of evidence-based internet interventions (eg, internet-based cognitive behavioral therapy) for the treatment of adults experiencing depressive symptoms across Europe [26]. In this study, we assessed the cost-effectiveness of the Super@ intervention as part of its implementation within the MasterMind project in a pilot site in Spain. The current analysis was performed using the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (MAFEIP) tool, developed for monitoring the financial sustainability of initiatives for promoting a healthy lifespan of European citizens [27,28]. Provided as a free-to-access tool for economic evaluations, MAFEIP has gained relevance over the years, and its usage is expanding, particularly within the European project landscape.

Methods

Overview of Study Design

As part of the MasterMind project for implementing an internet-based cognitive behavioral therapy for treating depression, we designed a pragmatic within-group trial to assess the cost-effectiveness of the intervention [29,30]. The evaluation framework applies the Model for Assessment of Telemedicine applications [31], which helped to define the data collection tools and instruments according to 3 levels of stakeholders involved within the implementation process: (1) patients, (2) professionals, and (3) organizations.

This analysis corresponds to the experience of the MasterMind project in the BSA (*Badalona Serveis Assistencials*) consortium, implemented under a program named *Super@ tu depresión* ("Get over your depression"). The BSA consortium provides primary and specialized care to a catchment population of 433,175 inhabitants in the most densely populated suburban area of Barcelona and has a long tradition in integrated care and the adoption of digital health solutions [32-38]. The implementation and data collection process for the Badalona pilot site was carried out between March 2015 and June 2017. The outcomes and costs of the intervention were compared with those of usual care in previously published data from the same area [39].

The local study protocol was approved by the Ethics Committee of the Hospital *Germans Trias i Pujol* (reference PI-15-069), and all participants provided informed consent before entering the study.

Participants

Study candidates included health care recipients and were screened for eligibility after general practitioner referral in the primary care setting. All consecutive patients who visited their general practitioners during the study period and met the selection criteria were offered the opportunity to participate in the Super@ program. Patients included in the study were adults (ie, 18 years or older) diagnosed with mild, moderate, or severe major depressive disorder based on the Patient Health Questionnaire 9 (PHQ-9; with score cutoffs of 10, 15, and 20 for mild, moderate, and severe major depressive disorder, respectively), living in Badalona and who, according to their general practitioner, had a certain level of technological literacy and internet connection. The main exclusion criteria were having

XSL•FO
RenderX

comorbidities that may interfere with the treatment, having a nonpsychiatric disease that could explain depressive symptoms, receiving structured face-to-face psychological therapy at the time of inclusion, and reporting a high suicidal risk or ideation (item 9 of the PHQ-9). After checking all selection criteria and obtaining written informed consent, the general practitioner referred participants to the internet-based cognitive behavioral therapy service, provided a comprehensive explanation about the intervention, and enrolled participants in the platform, which automatically provided a username and a password to the participant.

Intervention

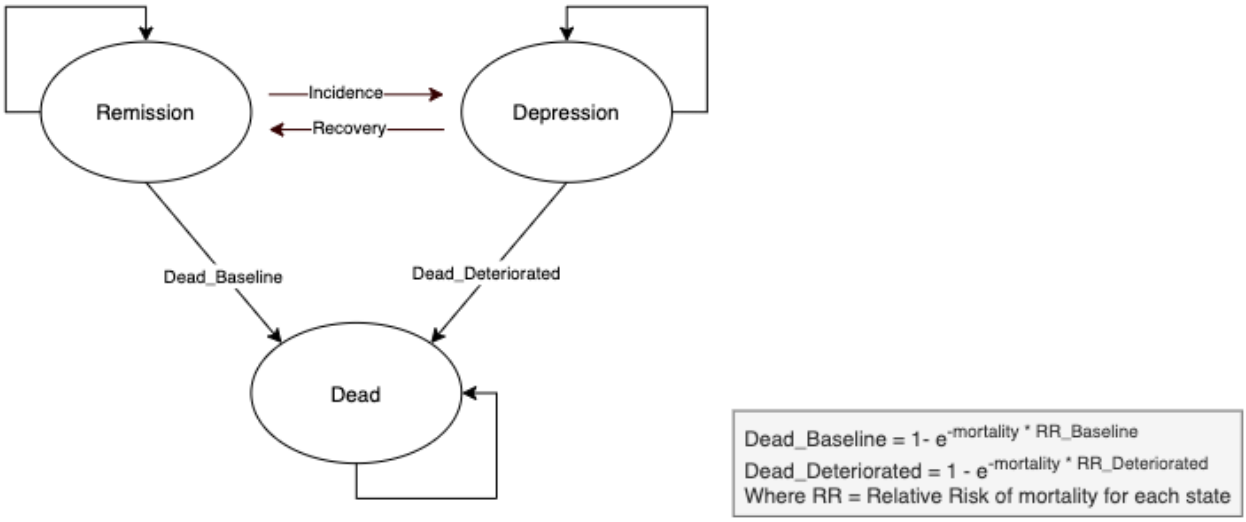
The Super@ program (Multimedia Appendix 1) consisted of 9 modules (8 regular and 1 extra) composed of videos, text content, and questionnaires to monitor the progression of symptoms and adherence to the intervention. Therapists provided guidance and project management within the BSA team to ensure patient follow-up and activation of the appropriate resources upon an increase of depressive symptoms. A project management team facilitated the project and its implementation process. Table S1 (Multimedia Appendix 1) summarizes the main activities performed in the project and the different professional profiles and teams involved in each. Intervention completion (ie, minimal adequate dose) was defined as engaging in a minimum of 3 modules of internet-based cognitive behavioral therapy.

Cost-Effectiveness Assessment

Model Structure, Transition Probabilities, and Utility Estimates

The cost-effectiveness of the Super@ program was assessed using the MAFEIP tool, which computes costs and utilities using a Markov model of health states and corresponding transition probabilities [40]. Based on previous economic evaluations of treatments for major depressive disorder, we defined a 3-state Markov model, with remission (PHQ-9 score <10), depression (PHQ-9 score ≥10), and death [41] (Figure 1). Transitions between the 3 states of the Markov model included recovery (ie, the probability of going from depression to remission) and relapse (ie, the probability of going from remission to depression); death was used as an absorbing state. The transition probabilities for the intervention group were calculated based on the changes between the health states at baseline and after the intervention. Given the lack of a control group, the corresponding probabilities for treatment as usual were obtained from a recent meta-analysis [41] assessing the usual care effects on major depressive disorder, which included 38 studies with pooled a remission rate (adjusted for publication bias) of 33% (95% CI 26%-40%). As suggested elsewhere [42], the risk of all-cause mortality was derived from life tables—in this case, the Human Mortality Database [43], which is stratified by gender and provides mortality rates at concrete years of age—and adjusted for depression [44].

Figure 1. Diagram of the 3-state Markov model of health states and transition probabilities designed for the Super@ intervention.



In accordance with standard guidelines for estimating quality-adjusted life-years in economic evaluations, the MAFEIP tool recommends computing utilities based on measures of health-related quality of life, preferably the EuroQoL 5-dimension (EQ-5D) questionnaire [45]. However, no estimates of health-related quality of life were collected during the assessment of the Super@ program. Alternatively, based on the relationship between utility scores derived from (including the EQ-5D score) and depression health states reported by Kolovos et al [46], we estimated the remission utility from the results of the PHQ-9 measure: the values proposed for 4 clinical categories of major depressive disorder severity were adapted to the 3-state model by estimating the weighted average of utilities of patients in the remission state (ie, PHQ-9 score <10) and those in the depression state (PHQ-9 score ≥10) (Multimedia Appendix 1).

Cost Estimate

The MAFEIP tool considers 3 types of costs: one-off costs, which represent the total cost incurred only once at the implementation point (ie, implementation support, training provision of professional staff, and cost of devices), health care costs (ie, health care resources consumption such as costs associated with the time spent by health care professionals on service provision, hospitalizations, pharmacy, etc), and societal

XSL•FO
RenderX

costs (ie, related to the time spent by either patients or informal caregivers such as the time spent using the technology or traveling to the hospital).

One-off costs were the main costs of implementing the Super@ services and included the support given to therapists to implement the intervention in their daily routine, the training of professional users, and the costs of development and adaptation of Super@ to the existing information and communication technology platform. Based on the annual gross salary of technical staff in Spain and the number of hours devoted to the project (ie, part-time 50%), we estimated €158 per patient (approximately US $190; an exchange rate of approximately €1 to US $1.20 is applicable at the time of publication). The costs of development and adaptation of Super@ to the existing information and communication technology platform were €237 per patient (US $285). Recurring costs, which included direct costs of each internet-based cognitive behavioral therapy session, amounted to €2439 (US $2927) per patient. For the control group, the typical situation is setting the one-off and recurrent costs at 0, because in most cases, the intervention would mean an additional investment.

Health care and societal costs were not collected in the MasterMind project. These costs were gathered from a previous study [39] that described the costs associated with major depressive disorder in the same area. Based on this study [39], we established the health care costs for patients in remission and depression as €451 and €826, respectively (US $542 and US $993). Correspondingly, the costs due to loss of labor productivity were €991 and €1842 for patients in remission and depression, respectively (US $1191 and US $2214). Health care and societal costs were assumed to be the same for the intervention and control groups.

In accordance with recommendations from local health technology assessment authorities in Spain, we applied a discount factor of 3% for both health care outcomes and costs [47]. The willingness-to-pay threshold was established at €30,000 per quality-adjusted life-year, the threshold most frequently used in Spain (equivalent to approximately $36,060).

The cycle length of the Markov model was set at 1 year (ie, the maximum allowed in the MAFEIP tool). Given the chronic nature of major depressive disorder [48], we established the number of cycles necessary to cover the time lapse between the average age of study participants (ie, 46 years) and a theoretical lifespan time of 100 years.

Analysis

The cost-effectiveness analysis was performed on the intention-to-treat sample, which included all participants who started at least 1 module of the treatment. The clinical and demographic characteristics of study participants were described with R software (version 3.5.3; The R Project) using the frequency percentage and the mean and standard deviation for categorical and quantitative variables, respectively. Variables of time were described as the median and interquartile range. The cost-effectiveness analysis was conducted using the MAFEIP tool including health states, transition probabilities, utility scores, and costs. All participants started on the state *depression* in the 3-state Markov model.

In addition to the base-case analysis, we conducted deterministic sensitivity analyses for 2 scenarios: reduction in session cost (up to 25%) associated with a lower professional-to-patient ratio expected for a scaling up of the intervention, and 0% to 5% discount in utilities, as recommended by local guidelines for economic evaluations [49]. Sensitivity analyses were nondiscounted.

Transition probabilities were computed using R software, whereas costs and utilities were calculated using a spreadsheet (Excel, version 2013; Microsoft Inc).

Results

Study Population and Intervention Conduct

Of the 253 patients recruited for the study, 229 participants (90.5%) started at least one module of the treatment (intention-to-treat sample), of whom 1 participant (0.4%) did not provide data on posttreatment status, and 81 participants (35.4%) did not complete treatment; therefore, 147 participants completed the treatment (Multimedia Appendix 1). All participants had been recruited during a clinical interview after referral by their general practitioner, and all completed the PHQ-9 questionnaire. Table 1 summarizes demographic and clinical characteristics of the intention-to-treat sample at baseline. Participants in the intention-to-treat sample remained under the Super@ program a median of 96 days (IQR 70-321); 147 participants (64.2%) were considered to have completed the study. At the end of the intervention, 98 participants (66.7%) had achieved the remission state. No adverse events related to the intervention or the major depressive disorder were reported.

XSL•FO
RenderX

Table 1. Clinical and demographic baseline characteristics of the participants who started the treatment.

Characteristic	Intention-to-treat sample (n=229)
Age (years), mean (SD)	46.40 (12.51)
Gender, n (%)	
Male	73 (31.9)
Female	156 (68.1)
Education, n (%)	
Primary	42 (18.3)
Secondary	100 (43.7)
Higher	78 (34.1)
Other	8 (3.5)
Not answered	1 (0.4)
Employment, n (%)	
Yes	169 (73.8)
No	58 (25.3)
Unknown	1 (0.4)
Not answered	1 (0.4)
Depressive episodes, n (%)	
Less than 4 weeks	10 (4.4)
Between 4 and 8 weeks	40 (17.5)
Between 8 and 12 weeks	65 (28.4)
Between 3 and 6 months	51 (22.3)
Between 6 months and 1 year	36 (15.7)
Between 1 year and 3 years	23 (10.0)
Between 3 and 5 years	2 (0.9)
Between 5 and 10 years	2 (0.9)
Antidepressant medication, n (%)	
Yes, for less than 1 month	7 (3.1)
Yes, for less than 2 months	44 (19.2)
Yes, for more than 2 months	74 (32.3)
No	104 (45.4)
Satisfaction with life[a], n (%)	
Preintervention	3.50 (1.16)
Postintervention	4.03 (1.28)
Satisfaction with mental health[a], n (%)	
Preintervention	3.23 (1.03)
Postintervention	3.98 (1.32)

[a]Assessed using a single-item question (How satisfied are you with your life as a whole today? or How satisfied are you with your mental health?) and rated on a 6-point scale (1=couldn't be worse, 2=displeased, 3=mostly dissatisfied, 4=mixed, 5=mostly satisfied, 6=pleased, 7=couldn't be better).

Study Parameters and Base Case Analysis

Table 2 summarizes the inputs of the cost-effectiveness analysis, including transition probabilities, costs, and utilities.

The Super@ program cost more than usual care from both health care and societal perspectives (Table 3).

XSL•FO
RenderX

Table 2. Inputs of the MAFEIP tool for computing the cost-effectiveness analysis.

Input	Control group	Intervention group (n=229)
Transition probabilities (%)		
Remission	14	0
Depression	29	48.53
Costs (€[a] per patient and year)		
One - off cost per patient	N/A[b]	395.26
Recurring cost per patient per year	N/A	2439
Health care cost		
Remission	451	451
Depression	826	826
Societal cost		
Remission	991	991
Depression	1842	1842
Utilities		
Remission	0.62	0.665
Depression	0.532	0.529
Relative risk of mortality		
Remission	1	1
Depression	1.68	1.68

[a]An exchange rate of approximately €1 to US $1.20 is applicable at the time of publication.

[b]N/A: not applicable.

Table 3. Incremental costs, effects, and cost-effectiveness ratio from health care and societal perspectives.

Perspective	Incremental cost (€[a])	Incremental effects (QALY[b])	Incremental cost-effectiveness ratio (€/QALY)
Health care perspective			
Discounted (3% for both costs and effects)	50,924.53	1.734	29,366.92
Nondiscounted	87,807.06	3.315	26,484.27
Societal perspective			
Discounted (3% for both costs and effects)	48,178.53	1.734	27,783.38
Nondiscounted	83,181.81	3.315	25,089.21

[a]An exchange rate of approximately €1 to US $1.20 is applicable at the time of publication.

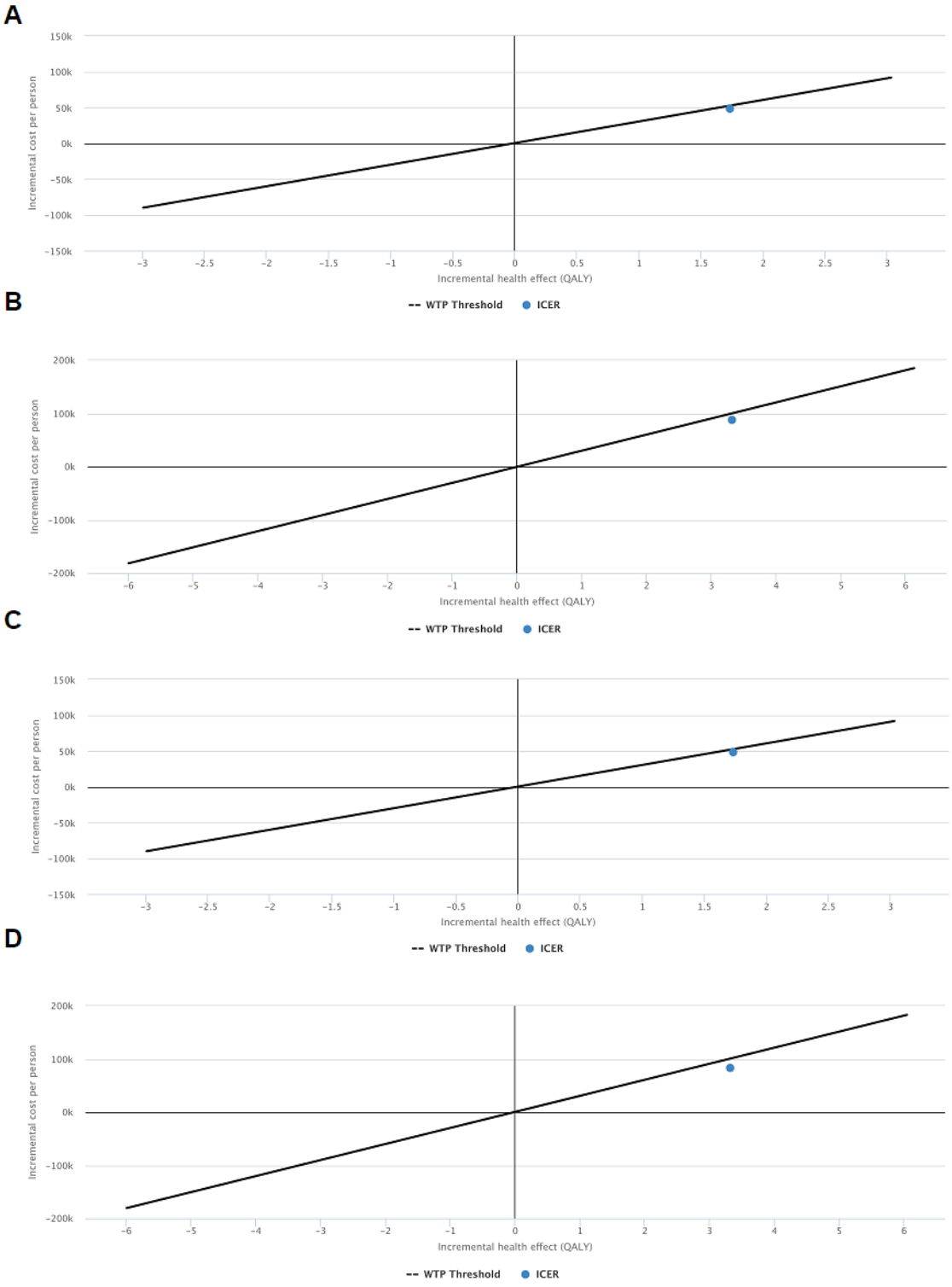
[b]QALY: quality-adjusted life-year.

The nondiscounted incremental cost-effectiveness ratios were below the willingness-to-pay threshold of €30,000 (Figure 2): €26,484 and €25,089 for health care and societal perspectives, respectively (US $31,833 and $30,162). The discounted incremental costs and effects were higher, although the incremental cost-effectiveness ratios remained below the willingness-to-pay threshold of €30,000.

In addition, we conducted a deterministic sensitivity analysis assuming that a greater number of participants to the program would results in a reduction of cost per session. A 25% reduction in the cost per session would reduce the incremental cost-effectiveness ratio from €26,484 to €19,623 (US $31,833 to $23,591) in the health care perspective analysis and from €25,089 to €18,228 (US $30,162 to $21,914) in the societal perspective analysis (Figure 3A and 3B). From the health care perspective (Figure 3C), the incremental cost-effectiveness ratio at the 5% discount in utility (worst-case scenario of the sensitivity analysis) was €71,041 (US $85,405). The corresponding intersection and lowest incremental cost-effectiveness ratio values for the societal perspective were 2.773 quality-adjusted life-years and €30,000 (Figure 3D).

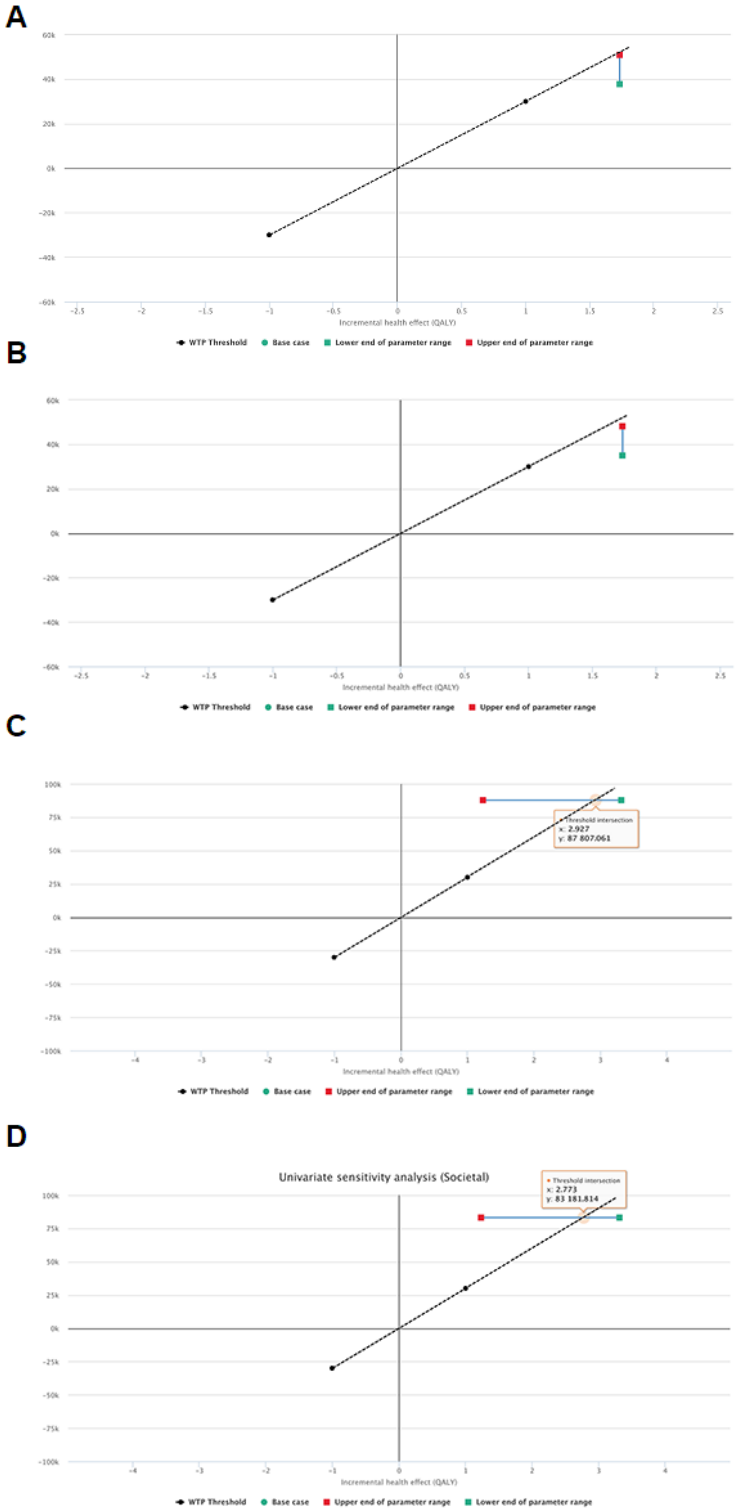
XSL•FO
RenderX

Figure 2. Cost-effectiveness plane of the Super@ intervention Healthcare perspective discounted (3% for both costs and health effects) (A) and non-discounted (B) analyses. Societal perspective discounted (3% for both costs and health effects) (C) and non-discounted (D) analyses. The solid line shows the 30,000 €/QALY willingness-to-pay threshold (equivalent to approximately US $36,060; an exchange rate of approximately €1 to US $1.20 is applicable at the time of publication). QALY: quality-adjusted life-year. WTP: willingness-to-pay.



XSL•FO
RenderX

Figure 3. Cost-effectiveness planes of sensitivity analyses. A reduction of up to 25% in cost per session (A and B for healthcare and societal perspectives, respectively), and 0% to 5% discount in effects (C and D for healthcare and societal perspectives, respectively). The dotted black line shows the 30,000 €/QALY willingness-to-pay threshold (equivalent to approximately US $36,060; an exchange rate of approximately €1 to US $1.20 is applicable at the time of publication). The solid green line shows the range of the incremental cost-effectiveness ratio, with the red and green squares indicating the range extremes for the worse (more costly or less effective) and best (less costly or more effective) scenario, respectively. QALY: quality-adjusted life-year. WTP: willingness-to-pay.



Discussion

In this analysis of the cost-effectiveness of an internet-based cognitive behavioral therapy intervention for mild or moderate major depressive disorder, we found that the intervention was more effective than treatment as usual, with incremental costs of €87,807 and €83,181 (nondiscounted from the health care and societal perspectives, respectively; US $105,561 and $99,999), according to costs reported for routine care of patients with mild-to-moderate major depressive disorder in our area

XSL•FO
RenderX

(Badalona, Spain). Despite the higher cost of the internet-based cognitive behavioral therapy intervention, it remained below the willingness-to-pay threshold of €30,000 typically used in Spain for making decisions in health care policies. According to the sensitivity analyses, the internet-based cognitive behavioral therapy would remain more expensive and more effective than treatment as usual in the onset of the cost reduction expected when scaling up the intervention (with the consequent decrease of the professional-to-patient ratio), with an incremental cost-effectiveness ratio below the willingness-to-pay threshold. When considering a 5% reduction in utility (ie, as suggested by local guidelines for economic evaluations), the intervention remained more effective than treatment as usual, although with an incremental cost-effectiveness ratio above the willingness-to-pay threshold.

In the last decade, many studies [50-53] have investigated the costs associated with internet-based cognitive behavioral therapy interventions, including therapies for major depressive disorder; however, most are based on descriptive approaches, which preclude drawing conclusions that can be used for making decisions on their implementation. More recently, Paganini et al [22] reviewed economic evaluations of internet-based cognitive behavioral therapy interventions for major depressive disorder that fulfilled preselected quality criteria, including the presence of comparator groups such as treatment as usual, another intervention, or wait-list controls. The case-mix of these interventions and heterogeneity of analyses precludes direct comparisons regarding the cost-effectiveness of each intervention. Nevertheless, they found that guided interventions (such as the Super@ program) tended to be more cost-effective than self-help ones, despite the higher cost associated with professional honoraria [22]. The incremental cost-effectiveness ratio of our intervention for the base-case health care perspective (€26,484 per quality-adjusted life-year) was in the lower zone of the wide range of values reported for guided interventions (ie, €19,616 [54] to €157,900 [55]; approximately US $19,616 and $189,825, respectively) and below that of unguided interventions (ie, €40,412 [56] to €178,700 [57]; approximately US $48,583 and $214,831, respectively).

Additionally, such studies [51,52] can only report on cost-effectiveness measures in controlled settings. Our study focused on the assessment of cost-effectiveness under real-world conditions free from biases possibly being introduced within efficacy studies such as a stricter application of protocolized procedures, eligibility criteria, and randomization [23-25]. Nevertheless, this approach resulted in some disadvantages, and our results should be interpreted with caution due to several limitations.

The lack of a comparator group has been considered among the limitations of economic evaluations of internet-based cognitive behavioral therapy interventions for major depressive disorder [22]. The pragmatic approach of our study, which took advantage of the implementation of the Super@ program by the local health care provider to assess its cost-effectiveness, precluded collecting treatment-as-usual data in parallel with those collected for the internet-based cognitive behavioral therapy intervention; however, the MAFEIP tool allowed us to rely upon literature for gathering these data. Of note, the source of cost-estimate data of treatment as usual for major depressive disorder corresponded to the same area in which the Super@ program was deployed [39]. Hence, the costs attributed to treatment as usual are expected to be similar to those we would have observed in a control group.

The MAFEIP tool also allowed us to bypass the unavailability of EQ-5D scores of health-related quality of life, a widely accepted measurement for computing utilities in cost-effectiveness analyses [45,58]. Other measures, such as disease severity scores, have been proposed as a proxy for health-related quality of life [59]. Taking advantage of the analysis by Kolovos et al, who established a relationship between health-related quality of life and PHQ-9 score for major depressive disorder severity [46], we computed the utility of the remission state of our 3-state Markov model using the PHQ-9 scores at the cutoff for minor depressive symptoms in the 5-state scale defined by the American Psychiatric Association [60] and the National Institute for Clinical Excellence [61].

Readers should take into consideration some limitations of the study design. First, the pragmatic approach constrained the number of participants to the implementation capacity, rather than the adequate sample size to achieve precision in our estimates. Second, like many other information and communication technology-based solutions, the success of an internet-based cognitive behavioral therapy intervention requires minimal technological literacy, which in our intervention was measured in an unstructured way at each general practitioners discretion. Technological literacy and keenness for the use of digital gadgets are expected to influence not only adherence but also the benefit that the patient may obtain from the intervention; the unstructured assessment of digital literacy may have introduced heterogeneity in the intervention outcomes. Third, the transferability of the results to other settings should be considered carefully. There are many reasons why cost-effectiveness analysis of health technologies may differ across jurisdictions and researchers and implementers should always refer to national guidelines in order to shed some light on the applicability of the results emerging from other contexts [62].

Our results suggest that the Super@ program provided benefits to patients at a cost that would allow its implementation in Spain, where interventions below €30,000 per quality-adjusted life-year are accepted. Costs associated with the intervention are expected to decrease in a scaling-up scenario; however, the sensitivity analysis of utility indicates that small reductions in effects would place the intervention at a nonacceptable incremental cost-effectiveness ratio based on the €30,000 threshold. Future studies should explore the patient profile that can benefit most from the intervention so that general practitioners have more information to target the therapy adequately.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Gerard Carot-Sans for providing medical writing support during the preparation of the manuscript. The authors would also like to thank the MasterMind project consortium partners for their contribution along the project.

The MasterMind project was partially funded under the Information and Communication Technology Policy Support Programme as part of the Competitiveness and Innovation Framework Programme by the European Community (grant agreement 621000).

The continuous development of the MAFEIP tool is funded by the European Commission's Horizon 2020 program under the projects WE4AHA (grant number 769705) and DigitalHealthEurope (grant number 826353).

The funders had no role in data collection or analysis, the decision to publish, or the preparation of the manuscript.

Authors' Contributions

JPJ was the principal investigator at the Badalona pilot site for the MasterMind project. JPJ and AE contributed to the study design and data collection. JPJ and AE conducted the statistical analyses. JPJ drafted the manuscript. AE and SK contributed to the different versions of the paper. FLV and FF supervised the study. All authors critically revised and approved the final version of the manuscript.

Conflicts of Interest

Author AE is employed by the Institute for health training online as research coordinator. All other authors declare no competing interests.

Multimedia Appendix 1
Supplementary file 1.
[PDF File (Adobe PDF File), 279 KB - jmir_v23i5e27410_app1.pdf]

References

1. Khan HTA. Population ageing in a globalized world: Risks and dilemmas? J Eval Clin Pract 2019 Oct;25(5):754-760. [doi: 10.1111/jep.13071] [Medline: 30485597]
2. Beard JR, Bloom DE. Towards a comprehensive public health response to population ageing. Lancet 2015 Feb 14;385(9968):658-661 [FREE Full text] [doi: 10.1016/S0140-6736(14)61461-6] [Medline: 25468151]
3. Bloom DE, Canning D, Fink G. Implications of population ageing for economic growth. Oxford Rev Econ Policy 2011 May 13;26(4):583-612. [doi: 10.1093/oxrep/grq038]
4. Hantrais L. Socio-demographic change, policy impacts and outcomes in social Europe. J Eur Soc Policy 2016 Dec 21;9(4):291-309. [doi: 10.1177/a010186]
5. Fougère M, Mérette M. Population ageing and economic growth in seven OECD countries. Econ Model 1999 Aug;16(3):411-427. [doi: 10.1016/s0264-9993(99)00008-5]
6. GBD 2017 DiseaseInjury IncidencePrevalence 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]
7. Bruce ML, Sirey JA. Integrated care for depression in older primary care patients. Can J Psychiatry 2018 Jul;63(7):439-446 [FREE Full text] [doi: 10.1177/0706743718760292] [Medline: 29495883]
8. Polyakova M, Sonnabend N, Sander C, Mergl R, Schroeter M, Schroeder J, et al. Prevalence of minor depression in elderly persons with and without mild cognitive impairment: a systematic review. J Affect Disord 2014 Jan;152-154:28-38. [doi: 10.1016/j.jad.2013.09.016] [Medline: 24103852]
9. 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]
10. Alexopoulos GS. Depression in the elderly. Lancet 2005;365(9475):1961-1970. [doi: 10.1016/S0140-6736(05)66665-2] [Medline: 15936426]
11. Husain SF, Yu R, Tang T, Tam WW, Tran B, Quek TT, et al. Validating a functional near-infrared spectroscopy diagnostic paradigm for major depressive disorder. Sci Rep 2020 Jun 16;10(1):9740 [FREE Full text] [doi: 10.1038/s41598-020-66784-2] [Medline: 32546704]
12. Ho R, Chua A, Tran B, Choo C, Husain S, Vu G, et al. Factors associated with the risk of developing coronary artery disease in medicated patients with major depressive disorder. Int J Environ Res Public Health 2018 Sep 21;15(10):2073 [FREE Full text] [doi: 10.3390/ijerph15102073] [Medline: 30248896]

XSL•FO
RenderX

13. Vu HTT, Nguyen TX, Nguyen HTT, Le TA, Nguyen TN, Nguyen AT, et al. Depressive symptoms among elderly diabetic patients in Vietnam. Diabetes Metab Syndr Obes 2018;11:659-665 [FREE Full text] [doi: 10.2147/DMSO.S179071] [Medline: 30425543]
14. Ng A, Tam WW, Zhang MW, Ho CS, Husain SF, McIntyre RS, et al. IL-1β, IL-6, TNF- α and CRP in elderly patients with depression or Alzheimer's disease: systematic review and meta-analysis. Sci Rep 2018 Aug 13;8(1):12050 [FREE Full text] [doi: 10.1038/s41598-018-30487-6] [Medline: 30104698]
15. Cuijpers P, Karyotaki E, de Wit L, Ebert D. The effects of fifteen evidence-supported therapies for adult depression:a meta-analytic review. Psychother Res 2020 Mar;30(3):279-293. [doi: 10.1080/10503307.2019.1649732] [Medline: 31394976]
16. Wittchen H, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Sep;21(9):655-679. [doi: 10.1016/j.euroneuro.2011.07.018] [Medline: 21896369]
17. Mack S, Jacobi F, Gerschler A, Strehle J, Höfler M, Busch MA, et al. Self-reported utilization of mental health services in the adult German population--evidence for unmet needs? results of the DEGS1-mental health module. Int J Methods Psychiatr Res 2014 Sep;23(3):289-303 [FREE Full text] [doi: 10.1002/mpr.1438] [Medline: 24687693]
18. Ho R, Mak K, Chua A, Ho C, Mak A. The effect of severity of depressive disorder on economic burden in a university hospital in Singapore. Expert Rev Pharmacoecon Outcomes Res 2013 Aug;13(4):549-559 [FREE Full text] [doi: 10.1586/14737167.2013.815409] [Medline: 23977979]
19. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
20. 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]
21. Zhang M, Ho R. Moodle: the cost effective solution for internet cognitive behavioral therapy (I-CBT) interventions. Technol Health Care 2017;25(1):163-165. [doi: 10.3233/THC-161261] [Medline: 27689560]
22. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 01;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
23. Wells KB. Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. Am J Psychiatry 1999 Jan;156(1):5-10. [doi: 10.1176/ajp.156.1.5] [Medline: 9892291]
24. Rothwell P. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
25. Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014 Jan 02;5:e45 [FREE Full text] [doi: 10.1038/ctg.2013.13] [Medline: 24384867]
26. Management of mental health disorders through advanced technology and services – telehealth for the mind. MasterMind Project. URL: https://mastermind-project.eu/ [accessed 2021-05-07]
27. Monitoring and assessment framework for the European innovation partnership on active and healthy ageing. European Commission. URL: https://www.mafeip.eu/ [accessed 2021-05-07]
28. Boehler CE, de Graaf G, Steuten L, Yang Y, Abadie F. Development of a web-based tool for the assessment of health and economic outcomes of the European Innovation Partnership on Active and Healthy Ageing (EIP on AHA). BMC Med Inform Decis Mak 2015 Sep 11;15(S3):1-1 [FREE Full text] [doi: 10.1186/1472-6947-15-s3-s4]
29. Campbell DT, Stanley JC. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally & Company; 2015.
30. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: he study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
31. Kidholm K, Ekeland A, Jensen L, Rasmussen J, Pedersen C, Bowes A, et al. A model for assessment of telemedicine applications: MAST. Int J Technol Assess Health Care 2012 Jan 23;28(1):44-51 [FREE Full text] [doi: 10.1017/s0266462311000638]
32. Baltaxe E, Cano I, Herranz C, Barberan-Garcia A, Hernandez C, Alonso A, et al. Evaluation of integrated care services in Catalonia: population-based and service-based real-life deployment protocols. BMC Health Serv Res 2019 Jun 11;19(1):370 [FREE Full text] [doi: 10.1186/s12913-019-4174-2] [Medline: 31185997]
33. Dueñas-Espín I, Vela E, Pauws S, Bescos C, Cano I, Cleries M, et al. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. BMJ Open 2016 Apr 15;6(4):e010301 [FREE Full text] [doi: 10.1136/bmjopen-2015-010301] [Medline: 27084274]
34. Rossi Mori A, Albano V, Piera Jimenez J. Badalona story: integrating the integration initiatives. Int J Integr Care 2017 Oct 17;17(5):315. [doi: 10.5334/ijic.3632]
35. RossiMori A, Piera J, Albano V, Mercurio G. A systematic analysis of the multi-annual journey of Badalona towards integrated care. Int J Integr Care 2019 Aug 08;19(4):344. [doi: 10.5334/ijic.s3344]
36. Moharra M, Vela E, Dueñas-Espín I, Pauws S, Bescos C, Cano I, et al. Health risk assessment and stratification in an integrated care scenario. Int J Integr Care 2016 Dec 16;16(6):322. [doi: 10.5334/ijic.2870]

XSL•FO
RenderX

37. Piera-Jiménez J, Winters M, Broers E, Valero-Bover D, Habibovic M, Widdershoven JWMG, et al. Changing the health behavior of patients with cardiovascular disease through an electronic health intervention in three different countries: cost-effectiveness study in the do cardiac health: advanced new generation ecosystem (Do CHANGE) 2 randomized controlled trial. J Med Internet Res 2020 Jul 28;22(7):e17351 [FREE Full text] [doi: 10.2196/17351] [Medline: 32720908]
38. Piera-Jiménez J, Daugbjerg S, Stafylas P, Meyer I, Müller S, Lewis L, et al. BeyondSilos, a telehealth-enhanced integrated care model in the domiciliary setting for older patients: observational prospective cohort study for effectiveness and cost-effectiveness assessments. JMIR Med Inform 2020 Oct 06;8(10):e20938 [FREE Full text] [doi: 10.2196/20938] [Medline: 33021490]
39. Sicras-Mainar A, Mauriño J, Cordero L, Blanca-Tamayo M, Navarro-Artieda R. [Costs and associated factors with optimal and suboptimal responses to the treatment of major depressive disorder]. Aten Primaria 2012 Nov;44(11):667-675 [FREE Full text] [doi: 10.1016/j.aprim.2012.04.007] [Medline: 22789772]
40. Weinstein M, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, ISPOR Task Force on Good Research Practices--Modeling Studies. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health 2003;6(1):9-17 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.00234.x] [Medline: 12535234]
41. Kolovos S, van Tulder MW, Cuijpers P, Prigent A, Chevreul K, Riper H, et al. The effect of treatment as usual on major depressive disorder: a meta-analysis. J Affect Disord 2017 Mar 01;210:72-81. [doi: 10.1016/j.jad.2016.12.013] [Medline: 28013125]
42. Sonnenberg F, Beck J. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]
43. Jdanov D, Jasilionis D, Shkolnikov V, Barbieri M. Human mortality database. In: Gu D, Dupre M, editors. Encyclopedia of Gerontology and Population Aging. Cham: Springer; 2019.
44. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014 Apr;171(4):453-462. [doi: 10.1176/appi.ajp.2013.13030325] [Medline: 24434956]
45. NICE. Position statement on EQ-5D-5L valuation set. Pharmaco Econ Outcomes News 2017 Aug 26;785(1):7-7 [FREE Full text] [doi: 10.1007/s40274-017-4257-4]
46. Kolovos S, Bosmans J, van Dongen JM, van Esveld B, Magai D, van Straten A, et al. Utility scores for different health states related to depression: individual participant data analysis. Qual Life Res 2017 Jul;26(7):1649-1658 [FREE Full text] [doi: 10.1007/s11136-017-1536-2] [Medline: 28260149]
47. López-Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, et al. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ 2010 Oct 20;11(5):513-520. [doi: 10.1007/s10198-010-0244-4] [Medline: 20405159]
48. Kolovos S, Bosmans JE, Riper H, Chevreul K, Coupé VMH, van Tulder MW. Model-based economic evaluation of treatments for depression: a systematic literature review. Pharmacoecon Open 2017 Feb 7;1(3):149-165. [doi: 10.1007/s41669-017-0014-7]
49. Mora-Ripoll R, Gilabert-Perramon A, Oliva-Moreno J, Puig-Junoy J. Guidance for economic evaluation and budget impact analysis for pharmaceuticals in Catalonia (Spain). Value Health 2014 Nov;17(7):A447 [FREE Full text] [doi: 10.1016/j.jval.2014.08.1193] [Medline: 27201217]
50. Tate D, Finkelstein E, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]
51. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
52. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Aug 03;45(16):3357-3376. [doi: 10.1017/s0033291715001427]
53. Arnberg F, Linton S, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
54. Hollinghurst S, Peters T, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct;197(4):297-304. [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
55. Kolovos S, Kenter RM, Bosmans JE, Beekman AT, Cuijpers P, Kok RN, et al. Economic evaluation of internet-based problem-solving guided self-help treatment in comparison with enhanced usual care for depressed outpatients waiting for face-to-face treatment: A randomized controlled trial. J Affect Disord 2016 Aug;200:284-292. [doi: 10.1016/j.jad.2016.04.025] [Medline: 27155071]
56. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. Psychol Med 2013 Jun 24;44(4):741-752. [doi: 10.1017/s0033291713001323]

XSL•FO
RenderX

57. Gerhards S, de Graaf LE, Jacobs L, Severens J, Huibers M, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318. [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
58. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value Health 2013 Jan;16(1):202-210 [FREE Full text] [doi: 10.1016/j.jval.2012.10.010] [Medline: 23337232]
59. Pickard A, Knight S. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care 2005 May;43(5):493-499 [FREE Full text] [doi: 10.1097/01.mlr.0000160419.27642.a8] [Medline: 15838415]
60. Rush J, First MB, Blacker D. Handbook of Psychiatric Measures. USA: American Psychiatric Association; May 2002:639-640.
61. National Collaborating Centre for Mental Health. Depression in Adults with a Chronic Physical Health Problem. Leicester, UK: National Collaborating Centre for Mental Health (UK); 2010.
62. Drummond M, Barbieri M, Cook J, Glick H, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value Health 2009 Jun;12(4):409-418 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00489.x] [Medline: 19900249]

Abbreviations

BSA: Badalona Serveis Assistencials
EQ-5D: EuroQoL 5-dimension questionnaire
MAFEIP: Monitoring and Assessment Framework for the European Innovation Partnership (on Active and Healthy Ageing).
PHQ-9: Patient Health Questionnaire–9

Edited by G Eysenbach; submitted 24.01.21; peer-reviewed by R Ho, V Strotbaum; comments to author 12.02.21; revised version received 18.02.21; accepted 11.04.21; published 11.05.21.

Please cite as:
Piera-Jiménez J, Etzelmueller A, Kolovos S, Folkvord F, Lupiáñez-Villanueva F
Guided Internet-Based Cognitive Behavioral Therapy for Depression: Implementation Cost-Effectiveness Study
J Med Internet Res 2021;23(5):e27410
URL: https://www.jmir.org/2021/5/e27410
doi:10.2196/27410
PMID:33973857

©Jordi Piera-Jiménez, Anne Etzelmueller, Spyros Kolovos, Frans Folkvord, Francisco Lupiáñez-Villanueva. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.05.2021. 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 Nutritional Interventions for Middle-Aged and Older Adults: Systematic Review and Meta-analysis

Caroline Robert[1,2], BSc, MSc; Mojisola Erdt[1,3], Dipl-Inform, PhD; James Lee[1], BSc; Yuanyuan Cao[1], LLB, JM, PhD; Nurhazimah Binte Naharudin[1,4], BSc; Yin-Leng Theng[1], BSc, MSc, PhD

[1]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[2]Department of Pharmacology, National University of Singapore, Singapore, Singapore

[3]Institute for Infocomm Research, A*STAR, Singapore, Singapore

[4]Lee Kuan Yew Centre for Innovative Cities, Singapore University of Technology and Design, Singapore, Singapore

Corresponding Author:
Caroline Robert, BSc, MSc
Wee Kim Wee School of Communication and Information
Nanyang Technological University
31 Nanyang Link
Singapore, 637331
Singapore
Phone: 65 69083444
Email: carolinerobertjong@gmail.com

Abstract

Background: The risk of development of chronic diseases related to poor nutrition increases with age. In the face of an aging population, it is important for health care sectors to find solutions in delivering health services efficiently and effectively to middle-aged and older adults.

Objective: The aim of this systematic review and meta-analysis was to consolidate the literature that reported the effectiveness of eHealth apps in delivering nutritional interventions for middle-aged and older adults.

Methods: A literature search from five databases (PubMed, CINAHL, Cochrane, Web of Science, and Global Health) from the past 5 years was performed. Studies were selected for inclusion that used eHealth to deliver nutritional interventions to adults aged 40 years and above, and reported health and behavioral outcomes. Two independent reviewers searched for research articles and assessed the eligibility of studies to be included in the review. A third reviewer resolved disagreements on study inclusion. We also assessed the quality of the included studies using the CONSORT 2010 checklist.

Results: A total of 70 studies were included for analysis. The study quality ranged from 44% to 85%. The most commonly used eHealth intervention type was mobile apps (22/70, 31%). The majority of studies (62/70, 89%) provided multicomponent health interventions, which aimed to improve nutrition and other health behaviors (eg, exercise, smoking cessation, medication adherence). Meta-analysis results indicated high and significant heterogeneity; hence, conclusions based on these results should be considered with caution. Nonetheless, the results generally showed that eHealth interventions improved anthropometric and clinical outcomes, but not behavioral outcomes such as fruit and vegetable consumption.

Conclusions: The use of eHealth apps to deliver health interventions has been increasing in recent years, and these apps have the potential to deliver health services to a larger group of people. Our findings showed that the effectiveness of eHealth apps to deliver health interventions for middle-aged to older adults was supported by the improvement of anthropometric and clinical outcomes. Future work could aim to develop research frameworks in administering eHealth interventions to address heterogeneity in this field of research.

(J Med Internet Res 2021;23(5):e15649) doi:10.2196/15649

KEYWORDS

eHealth; mHealth; nutritional intervention; nutrition apps; middle-aged; older adults; systematic review; meta-analysis

XSL•FO
RenderX

Introduction

The world is aging rapidly as people live longer and fertility rates decline. An aging population poses challenges to society as it is accompanied by a declining labor force and an increase in government spending in health care, thereby increasing the burden of health care management and delivery [1]. Moreover, the risk of developing chronic diseases such as hypertension, diabetes, and coronary heart diseases increases with age [2]. In fact, studies have shown that the risk of developing chronic diseases linked to a poor diet not only increases with age [3,4], but the onset of chronic diseases was also found to increase rapidly among the middle-aged population [5]. To delay the onset of these chronic conditions that are often related to poor diet and nutrition in middle-aged and older adults, it is important to improve diet and nutrition [6,7]. The need for an adequate, healthy, and well-balanced diet is thus essential, not only for the management of chronic diseases but also for their prevention [8,9].

To support the management of a healthy diet and lifestyle, many have turned to eHealth services. eHealth is an emerging field in health care that utilizes various technologies for the management and delivery of health services to users. eHealth encompasses mobile health (mHealth), which focuses on mobile devices and apps. eHealth technologies are invaluable in providing health care services that are personalized, timely, and efficient [10]. The use of eHealth technologies in the health care sector has been increasing in recent years [11,12]. In turn, there has been great interest in the research community for evaluating the efficacy of eHealth technologies in delivering health services and achieving positive health outcomes for users [11,13-15].

eHealth nutritional technologies typically aim to provide nutrition-related services that focus on aiding people with weight loss, in maintaining a healthy diet, and in supporting the self-management of nutrition-related chronic diseases and abating further regression of chronic conditions. Although studies have reported the success of eHealth nutritional interventions in targeting weight loss and promoting healthy eating habits [13,16-18], the study populations typically comprised mainly younger adults or adults. To our knowledge, the use of eHealth and evaluation of its effectiveness in providing nutrition-related services for older adults is much less frequently reported in the literature. A systematic review [19] did explore the use of eHealth among older adults aged 50 years and above, but with greater focus on the use of these technologies as a means for health promotion and primary prevention. The authors reported that the use of eHealth technologies was generally accepted among the older population in health promotion and primary prevention of diseases. Most of the studies reviewed involved older adults with weight problems, who used eHealth technologies that aimed to improve physical activity and diet. However, the authors did not report how successful the reviewed eHealth technologies were in improving the health of the study groups.

Therefore, the aims of this systematic review and meta-analysis were to consolidate the results of published research studies on the use of eHealth technologies for nutrition- and diet-related services for middle-aged and older adults. The onset of chronic diseases related to poor lifestyle habits increases with age [20], which could lead to higher stress and burden on health care systems in the face of an aging society [21]. Furthermore, eHealth was suggested to be advantageous in administering health care services to older adults owing to its efficiency and cost-effectiveness, but more research is needed to evaluate the effectiveness of eHealth on improving or maintaining health, which can strengthen evidence to support its use in administering health services in an effective and cost-effective manner [22,23]. Thus, the purpose of this systematic review and meta-analysis was to evaluate the effectiveness of eHealth nutritional interventions for the prevention and management of chronic diseases among middle-aged and older adults.

To address this aim, we established the following research questions: (1) Which eHealth technologies (eg, mobile phone, wearables) are most commonly used in eHealth nutritional interventions for middle-aged and older adults? (2) What are the common types of eHealth features provided by eHealth nutritional technologies? (3) Compared to non-eHealth interventions/standard care, does eHealth result in improvements to health and behavioral outcomes related to nutrition and diet? (4) Does the duration of eHealth interventions lead to better overall improvements in health and behavioral outcomes related to nutrition?

We define eHealth nutritional interventions as those using technologies such as mobile devices, telephones, wearables, sensors, and mobile and web-based apps to support users in achieving nutritional-related outcomes such as weight reduction, or changes in dietary intake or behavior. This support could be in the form of a reminder system, coaching calls, sharing of educational content, or sending motivational messages with the primary aim to support users in various activities such as setting and achieving nutritional health goals, recording dietary behavior, monitoring food intake, regulating eating habits, or tracking physical activity.

We performed a systematic review to consolidate the literature on the use of eHealth technologies in providing nutrition-related services to answer research questions 1 and 2. Additionally, meta-analyses were performed to evaluate the effectiveness of eHealth nutritional interventions in improving outcomes, addressing research questions 3 and 4. The strengths and limitations of employing such eHealth nutritional interventions in older adults are also discussed to provide considerations for future research in developing and implementing eHealth nutritional technologies.

Methods

Literature Search

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015 (PRISMA-P 2015) guidelines [24]. Systematic searches were performed in May 2020 using five databases: PubMed/Medline, CINAHL, Cochrane, Web of Science, and Global Health. We considered that these databases are sufficiently extensive to cover the literature on eHealth,

XSL•FO
RenderX

mHealth, public health, and health systems. In addition, manual searches for relevant studies were also performed from the reference lists of retrieved articles and directly in Google Scholar. Systematic searches were limited to the English language literature, human research, and year of publication between 2014 and 2019. We limited the search to articles published in the last 5 instead of 10 years, as technology is rapidly outdated and findings published more than 5 years ago might not be up to date or of high relevance to the current and future context of eHealth nutritional interventions. We identified the search categories based on the purpose of the systematic review. The search term "eHealth" was selected as it is the type of intervention of interest to our research questions, "nutrition" was selected since the purpose of the eHealth interventions of interest is to improve nutrition, and the search terms "middle-aged" and "elderly" were selected as these were our target populations. These search terms were entered in the Cochrane Library medical subject heading (MeSH) browser, and the relevant MeSH terms and synonyms were selected. MeSH terms represent a controlled vocabulary thesaurus maintained by the National Library of Medicine, which are used to index research articles for the PubMed and Medline databases. As shown in Table 1, the selected MeSH terms were used to perform full-text literature searches in the respective databases. Additionally, we used Boolean operators (eg, AND, OR, NOT) in our search strategy to provide a narrower and more productive search. More detailed information about the search strategies used for each database are presented in Multimedia Appendix 1.

Table 1. Medical subject heading (MeSH) search terms used for the literature search.

Search terms	MeSH terms
eHealth	eHealth OR telemedicine OR telehealth OR mhealth OR mobile health
Nutrition	Nutrition, diet, food, eating, food intake, ingestion, diet habit
Middle-aged and elderly	middle aged OR aged OR "aged, 80 and over" OR elderly

Inclusion Criteria and Selection of Studies

Studies were selected to be included in the systematic review based on the following inclusion criteria: (1) used a form of eHealth nutritional intervention for disease prevention and/or management; (2) involved adults aged 40 years and above; and (3) reported data regarding anthropometric measures (eg, weight, BMI, blood pressure readings), dietary behaviors, and other health outcomes (eg, self-report of physical and mental well-being).

Based on our definition, an eHealth nutritional intervention should encompass technologies that aid participants in losing weight, in maintaining a healthy diet, and in supporting the self-management of nutrition-related chronic diseases and abating further regression of chronic conditions. The use of eHealth to improve diet and nutrition among older adults could either be the primary or secondary aim of the intervention. Likewise, the intervention could be a stand-alone nutrition-only intervention or part of a multicomponent intervention such as those that aimed to improve other health behaviors (eg, physical activity).

We compiled a comprehensive list of outcome measures from the literature pertaining to nutrition, diet, and health outcomes that we could expect to be reported as the results of eHealth interventions, including: (1) anthropometric outcomes such as weight, BMI, waist circumference, hip circumference, body adiposity, and waist-to-hip ratio; (2) clinical outcomes such as liver enzyme and lipid profile, cholesterol level, blood pressure level, fasting blood glucose level, hemoglobin A_{1c} (HbA_{1c}) level, urinary sodium level, triglycerides level, fat mass or body fat, pulse pressure, insulin level, C-reactive protein, alkaline phosphatase, total white blood cell, alanine aminotransferase/aspartate aminotransferase, and serum creatinine; (3) behavioral outcomes such as dietary attitudes, dietary behavior, adherence to a Mediterranean diet, adherence to health behaviors, physical activity level, nutritional status, low salt content of purchased foods, low saturated fat and energy content of purchased foods, and the Framingham Heart Study cardiovascular disease (CVD) risk score [25,26]; (4) educational outcomes such as nutritional knowledge and health literacy; and (5) other outcomes such as psychological outcomes, quality of life, and app usage.

A study was excluded from the review if it (1) implemented a nonexperimental study design (eg, observational and case studies, study protocol); (2) was not a peer-reviewed research article (eg, conference proceedings, letters, commentaries); (3) did not report any health outcomes as aforementioned in the inclusion criteria; and (4) used an eHealth intervention as a follow-up intervention to observe maintenance of outcome changes from a previously administered health intervention that did not use eHealth.

Titles and abstracts of studies were first screened by an independent reviewer. During the first round of screening, studies that did not meet the eligibility criteria for selection were excluded from the review. Two independent researchers reviewed and screened the remaining studies based on the inclusion and exclusion criteria, and any disagreements were discussed and resolved with a third independent reviewer. Data and references were managed using EndNote software.

Data Extraction

The process of data extraction followed a standardized procedure as reported in previous systematic reviews [13,19,27]. This systematic review adhered to the guidelines proposed by Cochrane Handbook for Systematic Review of Interventions [28]. Data were extracted based on our research aims, and the inclusion and exclusion criteria. Multimedia Appendix 2 presents the relevant characteristics and data that were extracted from the included studies. With the aim to answer our research questions, we extracted data on the study design, participant

XSL•FO
RenderX

information, intervention description, outcome measures, and results. Study design comprised the study method or design, information regarding the groups involved in the study, duration of the intervention, measurement time points, and attrition rate. Participant information included sample size and sample selection criteria such as demographics, disease or health condition, mean age, and gender distribution. Interventions are described as a brief overview of the intervention and eHealth features. Outcome measures included the primary and secondary outcomes measured in the study. Results are described as an overview of the main primary and secondary findings of the study, including details on mean or percentage changes in outcomes, as well as significance levels where possible.

Quality Assessment

The quality of all included studies in the systematic review was assessed using the CONSORT (Consolidated Standards of Reporting Trials) 2010 checklist for reporting randomized controlled trials (RCTs) [29]. The CONSORT checklist includes a 25-point evaluation for RCTs. Although it is mostly used to assess RCTs, most of the criteria in the CONSORT checklist are applicable to other study designs. Quality assessment using the CONSORT checklist has also been adopted in other systematic reviews [13,27]. To assess quality, points were allocated to each criterion: 1 point was given to a fulfilled criterion, 0.5 points to a partially fulfilled criterion, 0 points to an unfulfilled criterion, and NA was indicated for criteria that were not applicable to the study. The quality of the systematic review was checked in accordance with the PRISMA-P 2015 checklist and reported in Multimedia Appendix 3.

Meta-analysis

A meta-analysis was performed using Cochrane Review Manager version 5.3 [30]. Only studies with an RCT design were included in the meta-analysis, as scores at baseline and postintervention are not independent [31]. Mean change scores from baseline to postintervention were included as the dataset for analysis. Postintervention results were treated as the scores or values collected at the end of the intervention period. SD values of the mean change scores were obtained as reported or calculated from the standard error or 95% CI if the study did not report the SD values [28]. Studies were excluded from the meta-analysis if they did not provide sufficient data for computing the mean change and SD values, or if the methodology and instruments used to measure the outcomes were not standardized and similar to the majority of studies in the meta-analysis. Our meta-analysis was performed using a random-effects model, with mean difference for outcomes that were measured and reported in a standardized manner (eg, weight in kilograms), and standardized mean difference for outcomes that were measured using different scales or measurements [26]. Publication bias for each outcome was assessed with funnel plots (Multimedia Appendix 4). Subgroup analyses based on the duration of the intervention period (ie, ≤3 months, 4-6 months, 7-12 months) were performed on outcomes that had at least three studies per subgroup to evaluate whether the intervention duration had differential effects on the outcomes.

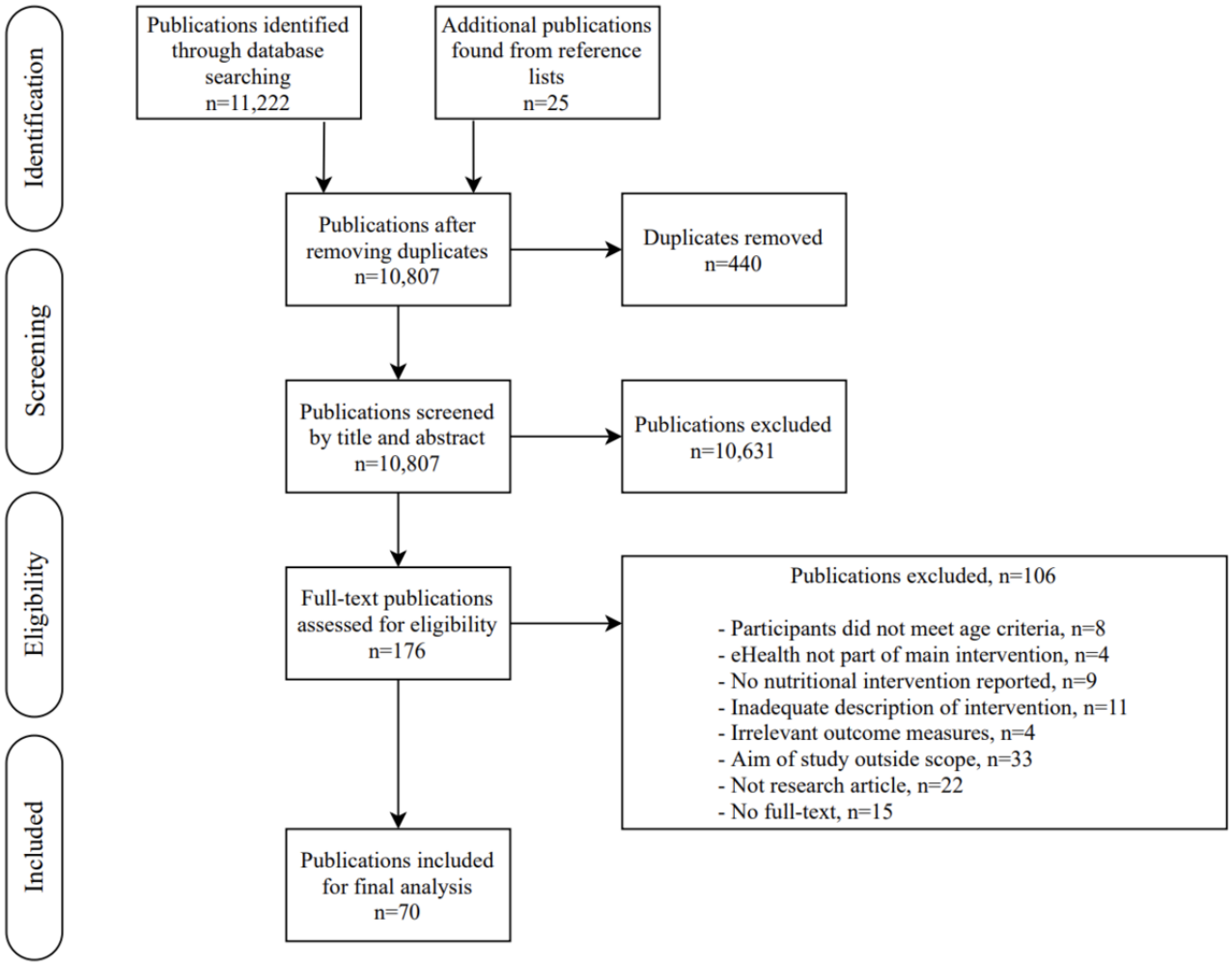
Results

Study Selection

Figure 1 presents the flowchart of the study selection process. A total of 11,247 studies were identified from database searches and additional reference lists. After the removal of 440 duplicates, 10,807 studies were screened by title and abstract. Among these, 176 full-text publications were considered to be potentially eligible for inclusion. We had several reasons for the exclusion of publications. Studies whose recruited participants were aged 39 years and below were excluded as they did not meet the age criterion of our target population. Publications were excluded that did not focus on eHealth as part of their main intervention. Publications were excluded that reported an eHealth intervention, but not related to nutrition. Publications that did not describe what the eHealth devices aimed to achieve in the nutritional intervention (eg, whether they comprised an educational component) were excluded due to the inadequate description of the intervention. Publications were also excluded for studies that did not measure any eHealth nutritional outcomes. Some publications were excluded because the study did not aim to administer an eHealth nutritional intervention to improve health outcomes of the participants but rather to, for example, compare methodologies or to evaluate the cost-effectiveness of eHealth programs, and were thus outside of the scope of this analysis. We only included full-text research articles. Of the full-text articles identified, 70 studies were finally included in the systematic review, having fulfilled our study inclusion criteria and having clearly indicated the effectiveness of implementing eHealth nutritional interventions on nutrition-related health outcomes.

XSL•FO
RenderX

Figure 1. Flowchart of publication selection procedure.



Study Characteristics

The characteristics of the included studies for the systematic review are presented in Multimedia Appendix 2. This systematic review adheres to the Cochrane guidelines for reporting characteristics of studies that were included in the review [28]. Thirteen studies were performed in the Asia-Pacific region [32-44], 3 studies were performed in the Middle East [45-47], 35 studies were performed in North America [48-82], 1 study was performed in South America [83], and 18 studies were performed in Europe [84-101]. The majority of the studies included were RCTs, with 38 studies that performed two-group RCTs [33-36,38-41,43,44,47-51,53,55,57-60,69,70,74-77,80,83,85,88,92,93,95-97,100,101] and 12 studies that performed three-group RCTs [45,46,54,56,61,79,86,87,90,91,94,99]. Nineteen studies involved a pre-post study design [32,37,42,52,63-68,71-73,78,81,82,84,89,98] and one study involved a pragmatic trial design [62].

Participant Information

The most common target population group was patients with chronic diseases. Of the 38 studies that recruited patients with chronic diseases, 25 recruited patients with diabetes mellitus or who were in a prediabetic state [32,40-44,46,47,50-52,55,60,62,65,66,70-72,75,82,83,86,90,93], 7 studies recruited patients with CVD [33,34,36,39,45,49,53], 1 study recruited patients with liver disease [48], 2 studies recruited patients with hypertension [61,69], and 3 studies recruited patients who reported any form of chronic disease [38,94,95]. Seventeen studies recruited patients with other clinical disorders such as individuals at risk of breast cancer [58], cancer survivors [64,92], those with obesity [57,59,63,67,68,73,76,77,89,101], and those who had obesity along with other health conditions [56,74,91,100]. The remaining 14 studies recruited healthy older adults [35,37,54,78-80,84,85,87,88,96-99]. With respect to gender, 47 studies recruited more female participants [32,34,37,44,48,52,54-59,62-66,68-72,76-78,80-87,89,91-93,95-97,99-101], 5 of which recruited only female participants [56,58,63,64,79]. One study recruited equal numbers of male and female participants [42]. The remaining 22 studies recruited more male participants for their intervention [33,35-41,43,45-47,49-51,53,60,61,67,74,90,94], 3 of which recruited only male participants [35,37,74]. Most of the studies (38/70, 54%) included participants with mean ages ranging from 50 to 59 years [32,33,39-43,45,46,48,51,53-55,58,60,61,64-66,69-71,73-75,79,81,82,84,86,88,90,92,94,96,97,100]. There were 18 studies that recruited participants with mean ages ranging from 40 to 49 years [34,35,37,47,57,59,63,68,72,76,77,80,85,87,89,91,99,101]. Eleven studies recruited participants with mean ages from 60 to 69 years [36,38,44,49,50,52,56,62, 67,83,93], and two studies recruited participants aged 70 and above [95,98]. Das et al [78] recruited participants with ages that ranged from

XSL•FO
RenderX

40 to 70 years and above. Recruitment ranged from 19 to 2305 participants, with reported attrition rates ranging from 0% to 67.4%.

eHealth Interventions

The purpose of using eHealth as a health intervention varies from the self-management of chronic diseases to the prevention or delay of their onset. The majority of the included studies (39/70, 56%) used eHealth interventions for the self-management of chronic conditions [32-34,36,39-50,53,55,60-62,65,66,69-71,74-76,78,79,82,83,86,90,93-95,100]. With healthy older adults as participants, 14 studies utilized eHealth as an intervention for the prevention of chronic diseases [35,37,54,77,80,84,85,87,88,96-99,101]. Additionally, 17 studies implemented eHealth interventions to support the progression of diseases [38,51,52,56-59,63,64,67,68,72,73,81,89,91,92].

The interventions were most commonly performed for a duration of 3-6 months. Twenty-nine studies performed the intervention for 3 months or less [32,34,36,40,41,45,46,57,59,60,63-65,70,74,76,78,80,82,86-88,91,93,96,97,101], 24 studies were performed for 4-6 months [33,37-39,42,44,47,48,51,53-56,58,61,68,73,75,81,84,92,98-100], 13 studies involved interventions for a duration of 7-12 months [35,43,52,62,67,69,70,84,85,89,90,94,95], and 3 studies performed the intervention for more than 12 months [66,70,79].

The studies included in this review reported the use of different technologies as their interventions, with the most common (22/70, 31%) being the use of mobile apps [36,40-42,44,47,53,57,63,64,66,68,75,77,84,89-91,96,97,99,101]. Twelve studies used a web-based app [34,43,62,65,67,69,70,78,79,87,92,100], 9 studies used phone calls [49,50,56,59,60,74,81,83,93], and 12 studies used wearable technology in conjunction with other eHealth technologies [37,38,52,55,58,76,79,80,86,88,94,100]. Two studies used an automated program [51,98] and the remaining studies used email [61,70,79], text messages [33,39,45,46,48,82], or videoconferencing [54,73,76,78,95]. As shown in Multimedia Appendix 2, the eHealth focused on nutritional intervention only in 8 studies [32,36,43,60,71,72,74,77,92,99,101]. The remaining studies administered an intervention on nutrition and other health behaviors such as exercise, smoking cessation, medication adherence, and behavioral change techniques for a better lifestyle.

Multimedia Appendix 5 presents the types of features implemented by the eHealth apps for the studies included in the systematic review. Most of the studies implemented several types of features in their eHealth app to deliver the intervention to participants. The most common feature of the eHealth interventions was the distribution of educational content for health behaviors. The second most common feature was to allow users to record their health behaviors such as dietary behavior, clinical and anthropometric data, or physical activity levels, and included health reports generated to assess their adherence or performance during the intervention. Thirty-six studies allowed participants to set their own health goals, which they would aim to achieve by the end of the intervention. Motivational messages would be sent to participants to further encourage them to adopt or maintain health behaviors, or to continue to adhere to the intervention. Fifteen studies implemented a reminder system in the mHealth app to send reminders to users to input their health data, and some included a point-based system whereby participants would receive incentives when they successfully attained a health goal.

Effectiveness of eHealth Interventions

A summary of the intervention findings that included anthropometric, clinical, behavioral, educational, and psychological outcomes is provided in Multimedia Appendix 6. In general, the studies found overall positive effects of the intervention in measured outcomes. The most commonly measured outcomes were weight (n=44), dietary behavior (n=30), BMI (n=27), HbA_{1c} levels (n=24), and physical activity levels (n=23). Among outcomes reported by five or more studies, adherence to health behaviors, Framingham CVD risk score, dietary behavior, weight, and BMI were those that showed the most common improvements in favor of the eHealth intervention group. Specifically, 5/7 (71%) of studies that reported adherence to health behaviors, 3/5 (60%) studies for Framingham CVD risk, 14/30 (47%) studies for dietary behavior, 20/44 (45%) studies for weight, and 11/27 studies (41%) for BMI showed improvements in favor of the intervention group. Multimedia Appendix 6 presents a summary of results for each included study at the latest measurement time point (ie, end of intervention or last follow-up), showing positive, negative, or no change to outcomes measured for all included studies in the intervention and control or comparison groups. Multimedia Appendix 7 presents a more comprehensive overview of the results for each study included in the systematic review. There are several cautionary notes about the results reported in some of the included studies. Bentley et al [86] did not perform inferential analyses of secondary outcomes (weight and HbA_{1c} level) due to a small sample size. Thus, positive results in favor of the eHealth intervention group should be interpreted with caution. Recio-Rodriguez et al [97] did not provide follow-up results on secondary outcomes (blood pressure, waist circumference, and BMI).

Meta-analysis

Summary of the meta-analysis findings are reported in Table S1 in Multimedia Appendix 8, including heterogeneity, mean differences and 95% CIs, as well as the significance of the intervention effect. The forest plots for each variable reported in Table S1 are shown in Figures S1-S12 in Multimedia Appendix 8. The *P* values for heterogeneity showed significance for nearly all variables included in the meta-analysis. This indicates significant clinical heterogeneity [102] among the studies, possibly due to variability in the eHealth nutritional interventions administered. The forest plots also show the diverse effects of the eHealth nutritional interventions on the outcomes reported in the meta-analysis. Nonetheless, we considered it to still be worthwhile to summarize the results quantitatively with a random-effects model [102]. The effects of eHealth interventions on weight, BMI, waist circumference, low-density lipoprotein-cholesterol, systolic blood pressure, and HbA_{1c} level favored the intervention group (all $P<.001$) as compared to the control group, whereas the effects on fruit and vegetable consumption favored the control group ($P=.01$). In

XSL•FO
RenderX

addition, the effects of the eHealth intervention on body fat, triglyceride level, Framingham CVD risk, and calorie intake favored neither the intervention nor the control group (all $P>.05$). Subgroup analyses based on the duration of the intervention period (ie, ≤3 months, 4-6 months, 7-12 months) on the outcomes of weight, waist circumference, and HbA_{1c} levels indicated high and significant heterogeneity for many variables (Figures S1, S3, and S7 in Multimedia Appendix 8); thus, the conclusions based on these results should be considered with caution.

eHealth Usage

The majority of the studies included in the systematic review neither reported participants' adherence to the eHealth app nor the participants' perceptions and satisfaction toward the eHealth intervention. Only two studies reported higher adherence to the mHealth interventions among the intervention group when compared to control groups [33,82,97]. A few studies reported that overall, the eHealth apps were well-received by participants, as they found it to be useful [36], easy to use [32,33,36,37,55,68,86], and were satisfied with their respective eHealth apps [33,68]. Interestingly, two studies found that older participants were more likely to use the eHealth intervention app more frequently than younger participants [88,90]. However, Fukuoka et al [55] raised the issue of decreasing adherence to the mHealth intervention due to technical issues with the mobile app and pedometer. Similarly, Mundi et al [68] found a decrease in mobile app usage over the study period, even though participants reported that they were generally satisfied with the mobile app.

Study Quality Assessed by the CONSORT 2010 Checklist

Multimedia Appendix 9 presents a detailed quality assessment of each study, with their raw and percentage scores indicated. The percentage scores for all included studies ranged between 44% and 85%. Overall, the quality for all included studies was judged to be "fair," with a mean percentage score of 65.85%. Two studies were assessed to have low quality [40,63] with percentage scores below 50%. Ten studies were ranked as fair with percentage scores between 50% and 59% [34,38,41,49,58,64,74,84,94,95]. Twenty-five studies with percentage scores ranging from 60% to 69% were assessed as good quality [32,35,39,45,52-54,56,57,60,63,65,67,70,71,83,85,86,89-93,96,98], and 18 studies were ranked as very good quality with a score of 70% and above [33,36,37,48,50,51,55,59,61,62,66,69,72,73,75,87,88,97].

Discussion

Principal Results

The findings of the studies included in this systematic review and meta-analysis provided varying evidence for the effectiveness of eHealth interventions in improving health and other related outcomes for the prevention and management of chronic diseases related to poor nutrition. Results from the systematic review demonstrated that eHealth interventions were highly successful in significantly improving adherence to health behavior, Framingham CVD risk, dietary behavior, weight, and BMI.

The results of the meta-analysis revealed overall positive within-group improvements in favor of the intervention group for anthropometric (ie, weight, BMI, waist circumference) and clinical (ie, low-density lipoprotein-cholesterol, systolic blood pressure, HbA_{1c} level) outcomes. No within-group improvements were found in fasting blood glucose, body fat, triglyceride level, Framingham CVD risk, and calorie intake. Subgroup analyses based on the intervention duration were performed for outcomes of weight, waist circumference, and HbA_{1c} levels. Regardless of the intervention duration, there were significant improvements in weight. However, the highest improvement in weight was found for interventions administered for 4-6 months. Similarly, interventions that were administered for exactly 4-6 months improved waist circumference and had the largest improvement on HbA_{1c} levels compared with other intervention durations. In other words, interventions offered for less than 4 months and for more than 6 months showed no significant improvements in waist circumference, and improvements in weight and HbA_{1c} level were lower as compared to those achieved when interventions were delivered for 4-6 months. This could most likely be due to the effectiveness of eHealth interventions often only being measurable after a minimum intervention duration. Moreover, studies with longer intervention durations could show less effectiveness possibly due to unsustainable rates of adherence and drops in compliance of the study participants over time. Nevertheless, interpretation of these results requires caution owing to the significant heterogeneity among studies included in the meta-analysis. With regard to behavioral outcomes, improvements in fruit and vegetable consumption were more apparent in the control group. However, caution is also needed in interpreting this finding as the meta-analysis only included three studies that measured fruit and vegetable consumption. Indeed, it is also important to consider that not all studies in the systematic review were included in the meta-analysis, as some studies reported insufficient data, or used very diverse instruments or methods to measure their outcomes. Therefore, caution is also required in interpreting the results related to the effectiveness of eHealth on health and behavioral outcomes.

Furthermore, discrepancies in the findings of the meta-analysis (ie, improvements in anthropometric and clinical outcomes, but not behavioral outcomes) suggest that the improvements found in eHealth intervention groups could be due to the fact that many studies not only focused on nutrition but also on other lifestyle behaviors. Moreover, several studies [32,33,36,57] found improvements in dietary behavior or adherence to health behavior, but did not find any improvements in anthropometric outcomes or clinical outcomes (eg, weight, BMI, blood pressure, cholesterol). This contrast in findings begs the question as to whether improvements in health outcomes were brought about by improvements in dietary behavior. Therefore, future studies should further evaluate the relationship between improvements in behavioral and health outcomes, and whether such relationships will lead to longer-lasting effects. This would help

XSL•FO
RenderX

validate the efficacy of the eHealth intervention, and its effectiveness in improving and maintaining health.

The implementation of eHealth technologies to deliver nutritional interventions provides great convenience and potential for health care systems, as participants are able to receive efficient, timely, and personalized health services. Previous research acknowledged the advantages of using eHealth in providing health services to older adults in a cost-effective and convenient manner [20,23]. Moreover, Kampmeijer et al [19] demonstrated that eHealth services are generally widely accepted among the older population. As these findings suggested that future work is needed to evaluate the effectiveness of eHealth on health outcomes, our current work bridges this gap in the literature by providing further evidence and considerations on the relevance of using eHealth for middle-aged and older populations.

Future Considerations

Due to the exploratory nature of our review, we included a considerably broad range of eHealth interventions, target outcomes, as well as population groups. To address concerns of heterogeneity found in the meta-analysis, future meta-analyses should focus on specific types of eHealth interventions (eg, mobile apps), specific primary outcomes (eg, weight or BMI), or specific population groups (eg, patients with hypertension). With heterogeneity addressed in this way, results and conclusions drawn upon these studies would then be able to offer insight into the efficacy of eHealth interventions on specific outcomes. Subgroup analyses could also be performed to evaluate if different intervention durations have differential effects on target outcomes.

The vast majority of the included studies used a multicomponent intervention, whereby the eHealth technologies were most commonly used to provide nutritional and physical exercise intervention. In fact, we found only six studies that solely delivered nutritional interventions, and reported mixed results on the efficacy of eHealth to participants' health outcomes [32,36,60,71,74,92]. The meta-analysis further showed that the intervention duration did not have a significant impact on health outcomes, suggesting that there could be other factors (eg, types of eHealth used or target population groups) that contributed to the effectiveness of such interventions. Having multicomponent interventions delivered through eHealth apps would understandably be more effective as compared to stand-alone nutritional interventions. Previous systematic reviews on mHealth app usage by adolescents and adults found similar favorable results for multicomponent interventions [13]. Moreover, for one to achieve a healthy and balanced lifestyle, it would be necessary to make positive changes to both diet and exercise simultaneously, rather than focusing on diet or exercise alone [103]. Thus, future developments of eHealth apps could consider having multicomponent intervention features to improve the health outcomes of users.

By and large, participants were reportedly satisfied with the eHealth technologies in helping them to achieve their health goals. However, a large number of studies did not systematically measure participants' adherence, perception, and satisfaction toward the eHealth technologies. Future studies on the efficacy of eHealth interventions should measure these aspects, as it is important to have a better understanding of participants' evaluation of the eHealth technologies from a qualitative and quantitative perspective. Otherwise, if the interventions provided are not widely accepted by participants, this might reduce their efficacy. Furthermore, if we consider the lack of a scientific theory and framework for app development [104], it would be imperative for future studies to consider participants' engagement and evaluation of eHealth technologies, which might effectively improve future app development.

Strengths and Limitations

One of the limitations of this systematic review was the search strategy employed to obtain relevant research articles. We searched for articles from five databases, which might have led us to miss relevant articles in the literature. However, we considered that the databases used for our search were sufficiently extensive in covering topics on eHealth and health services. We also manually searched articles that were not found in our selected databases. In addition, there are many ways to describe the search terms, which we might not have exhaustively included in our search strategy. However, we do consider the inclusion of MeSH terms, and general terms such as "telemedicine" and "nutrition" should suffice in comprising the main terms and concepts of interest. Nevertheless, the large number of articles identified, screened, and excluded does indicate that the search query could have been more specific. In addition, the use of Boolean operators (OR, AND) in our search query might have led to different interpretations by the various databases. Another limitation of our search strategy was the timeframe of searching for articles published in the last 5 years. We considered a 5-year timeframe as sufficient, as technologies used in eHealth tend to change and improve rapidly, and therefore become outdated in a short period. As such, studies using older technologies might already be outdated. We encountered important clinical heterogeneity in our meta-analysis; however, we still decided to report our findings in the appendices to illustrate that although these studies aimed to measure the same outcomes, there was high heterogeneity due to the diverse ways the interventions were administered (eg, mobile apps vs web-based tools). In addition, due to the different types of interventions administered across studies, the issue of clinical heterogeneity [102] makes it difficult to conclude consistent and convincing meta-analysis findings. Finally, we did not account for the quality of studies in our subgroup analyses. Variability in study quality may overestimate positive results, especially if studies with poorer quality are included in the meta-analysis. Nonetheless, the majority of our studies were ranked as being of good or very good quality; thus, we might expect less variability on the effect of study quality on subgroup analyses.

One strength of our study was that we had two independent reviewers performing the selection process and a third independent reviewer solving any disagreements in study inclusion. In addition, we followed the PRISMA-P 2015 guidelines for reporting a systematic review, and checked the quality of each included study using the CONSORT 2010 checklist in a standardized manner. We did consider additional ways to assess study quality, such as using the Cochrane risk

XSL•FO
RenderX

of bias tool [105], which assesses each study outcome. However, we found that due to the exploratory nature and scope of our research aims, we were considering a rather large number of studies (70 in total), each having multiple primary and secondary outcomes. This made it challenging to assess the risk of bias for each outcome across all studies. Although the CONSORT checklist does not explicitly assess risk of bias, it does score and assess whether studies have adequately designed, analyzed, interpreted, and reported their results and methods, which we believe is an acceptable and valid way of assessing study quality and risk of methodological bias [13,27].

Conclusion

We found varying evidence for the effectiveness of eHealth in providing nutrition-related interventions for middle-aged and older adults. Studies included in the systematic review generally reported positive support for the use of eHealth technologies in improving health and behavioral outcomes. Apps that delivered multicomponent interventions for improving nutrition and other health behaviors were more commonly used in health interventions as compared to stand-alone apps. The use of eHealth technologies has been increasing over the years, and has great potential in effectively delivering health services to a large group of people. Meta-analysis findings, although based on heterogeneous data and with quality limitations, generally demonstrated improvements in weight and BMI for eHealth users. To address heterogeneity in this field of research, future studies could look into developing a research framework or consensus in administering studies that involve eHealth interventions. Nevertheless, more research is needed to measure participants' engagement with the technologies, and to provide structural and scientific frameworks for the development of future apps, which can provide a better understanding of their effectiveness and encourage app adherence among users.

Acknowledgments

This research is supported by the National Research Foundation Singapore under its National Innovation Challenge on Active and Confident Ageing (grant MOH/NIC/CAHIG03/2016), administered by the Singapore Ministry of Health National Medical Research Council.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies used for each database search.
[DOCX File , 14 KB - jmir_v23i5e15649_app1.docx]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File , 31 KB - jmir_v23i5e15649_app2.docx]

Multimedia Appendix 3
PRISMA-P 2015 Checklist.
[DOCX File , 33 KB - jmir_v23i5e15649_app3.docx]

Multimedia Appendix 4
Funnel plots.
[DOCX File , 2178 KB - jmir_v23i5e15649_app4.docx]

Multimedia Appendix 5
Types of features implemented by eHealth apps.
[DOCX File , 21 KB - jmir_v23i5e15649_app5.docx]

Multimedia Appendix 6
Summary of findings.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i5e15649_app6.xlsx]

Multimedia Appendix 7
Comprehensive overview of studies included in the systematic review.
[DOCX File , 196 KB - jmir_v23i5e15649_app7.docx]

Multimedia Appendix 8

XSL•FO
RenderX

Summary of the meta-analysis results and forests plots.
[DOCX File , 2713 KB - jmir_v23i5e15649_app8.docx]

Multimedia Appendix 9
Study quality assessed by CONSORT 2010 Checklist.
[XLSX File (Microsoft Excel File), 235 KB - jmir_v23i5e15649_app9.xlsx]

References

1. WHO global strategy on people-centred and integrated health services. World Health Organization. 2015. URL: https://www.who.int/servicedeliverysafety/areas/people-centred-care/global-strategy/en/ [accessed 2021-05-10]
2. Denton FT, Spencer BG. Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. Can J Aging 2010 Mar;29(1):11-21. [doi: 10.1017/S0714980809990390] [Medline: 20202262]
3. Nishida C, Uauy R, Kumanyika S, Shetty P. The joint WHO/FAO expert consultation on diet, nutrition and the prevention of chronic diseases: process, product and policy implications. Public Health Nutr 2004 Feb 02;7(1A):245-250. [doi: 10.1079/phn2003592] [Medline: 14972063]
4. Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation WHO Technical Report Series, No. 916 (TRS 916). World Health Organization. 2003. URL: https://www.who.int/dietphysicalactivity/publications/trs916/en/ [accessed 2021-05-10]
5. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007 May 28;167(10):1068-1074. [doi: 10.1001/archinte.167.10.1068] [Medline: 17533210]
6. Bernstein M, Munoz N, Academy of Nutrition Dietetics. Position of the Academy of Nutrition and Dietetics: food and nutrition for older adults: promoting health and wellness. J Acad Nutr Diet 2012 Aug;112(8):1255-1277. [doi: 10.1016/j.jand.2012.06.015] [Medline: 22818734]
7. Meydani M. Nutrition interventions in aging and age-associated disease. Ann N Y Acad Sci 2001 Apr;928(1):226-235. [doi: 10.1111/j.1749-6632.2001.tb05652.x] [Medline: 11795514]
8. Marsman D, Belsky DW, Gregori D, Johnson MA, Low Dog T, Meydani S, et al. Healthy ageing: the natural consequences of good nutrition-a conference report. Eur J Nutr 2018 Jun;57(Suppl 2):15-34 [FREE Full text] [doi: 10.1007/s00394-018-1723-0] [Medline: 29799073]
9. Roberts CK, Barnard RJ. Effects of exercise and diet on chronic disease. J Appl Physiol 2005 Jan;98(1):3-30 [FREE Full text] [doi: 10.1152/japplphysiol.00852.2004] [Medline: 15591300]
10. Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the Internet age. J Med Internet Res 2001 Jun 7;3(2):e19 [FREE Full text] [doi: 10.2196/jmir.3.2.e19] [Medline: 11720961]
11. Hussain M, Al-Haiqi A, Zaidan A, Zaidan B, Kiah M, Anuar NB, et al. The landscape of research on smartphone medical apps: Coherent taxonomy, motivations, open challenges and recommendations. Comput Methods Programs Biomed 2015 Dec;122(3):393-408. [doi: 10.1016/j.cmpb.2015.08.015] [Medline: 26412009]
12. Silva BM, Rodrigues JJ, 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]
13. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
14. Wang J, Wang Y, Wei C, Yao N, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
15. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb 20;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
16. Allman-Farinelli M, Partridge SR, McGeechan K, Balestracci K, Hebden L, Wong A, et al. A mobile health lifestyle program for prevention of weight gain in young adults (TXT2BFiT): nine-month outcomes of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jun 22;4(2):e78 [FREE Full text] [doi: 10.2196/mhealth.5768] [Medline: 27335237]
17. Gilliland J, Sadler R, Clark A, O'Connor C, Milczarek M, Doherty S. Using a smartphone application to promote healthy dietary behaviours and local food consumption. Biomed Res Int 2015;2015:841368. [doi: 10.1155/2015/841368] [Medline: 26380298]
18. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth lifestyle program with telephone support (TXT2BFiT) to prevent unhealthy weight gain in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2015 Jun 15;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]

XSL•FO
RenderX

19. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
20. Atella V, Piano Mortari A, Kopinska J, Belotti F, Lapi F, Cricelli C, et al. Trends in age-related disease burden and healthcare utilization. Aging Cell 2019 Feb 29;18(1):e12861. [doi: 10.1111/acel.12861] [Medline: 30488641]
21. Marrero SL, Bloom DE, Adashi EY. Noncommunicable diseases: a global health crisis in a new world order. JAMA 2012 May 16;307(19):2037-2038. [doi: 10.1001/jama.2012.3546] [Medline: 22665101]
22. Arak P, Wójcik A. Transforming eHealth into a political and economic advantage (Polityka Insight). European Commission. 2017. URL: https://digital-strategy.ec.europa.eu/en/library/transforming-ehealth-political-and-economic-advantage [accessed 2021-05-10]
23. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: A systematic review. PLoS One 2018 Jun 13;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]
24. 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):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
25. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998 May 12;97(18):1837-1847. [doi: 10.1161/01.cir.97.18.1837] [Medline: 9603539]
26. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753. [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
27. 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 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
28. Higgins J, Wells G. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2019.
29. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c332-c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
30. Cochrane (RevMan) Version 5.2. Cochrane Training. Copenhagen: The Nordic Cochrane Centre; 2012. URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman [accessed 2021-05-10]
31. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2017 Aug;26(4):364-368 [FREE Full text] [doi: 10.1017/S2045796016000809] [Medline: 27790968]
32. Ahn Y, Bae J, Kim H. The development of a mobile u-Health program and evaluation for self-diet management for diabetic patients. Nutr Res Pract 2016 Jun;10(3):342-351 [FREE Full text] [doi: 10.4162/nrp.2016.10.3.342] [Medline: 27247732]
33. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
34. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018 Nov 19;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
35. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
36. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: Findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep 20;24(13):1435-1444. [doi: 10.1177/2047487317715713] [Medline: 28631933]
37. Gilson ND, Pavey TG, Wright OR, Vandelanotte C, Duncan MJ, Gomersall S, et al. The impact of an m-Health financial incentives program on the physical activity and diet of Australian truck drivers. BMC Public Health 2017 May 18;17(1):467 [FREE Full text] [doi: 10.1186/s12889-017-4380-y] [Medline: 28521767]
38. Lim S, Kang SM, Kim KM, Moon JH, Choi SH, Hwang H, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2016 Apr 5;53(2):189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]
39. Santo K, Hyun K, de Keizer L, Thiagalingam A, Hillis GS, Chalmers J, et al. The effects of a lifestyle-focused text-messaging intervention on adherence to dietary guideline recommendations in patients with coronary heart disease: an analysis of the TEXT ME study. Int J Behav Nutr Phys Act 2018 May 23;15(1):45 [FREE Full text] [doi: 10.1186/s12966-018-0677-1] [Medline: 29792202]

XSL•FO
RenderX

40. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
41. Zhou W, Chen M, Yuan J, Sun Y. Welltang - A smart phone-based diabetes management application - Improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]
42. Koot D, Goh PSC, Lim RSM, Tian Y, Yau TY, Tan NC, et al. A mobile lifestyle management program (GlycoLeap) for people with type 2 diabetes: single-arm feasibility study. JMIR Mhealth Uhealth 2019 May 24;7(5):e12965 [FREE Full text] [doi: 10.2196/12965] [Medline: 31127720]
43. Ramadas A, Chan CKY, Oldenburg B, Hussein Z, Quek KF. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes: changes in health cognitions and glycemic control. BMC Public Health 2018 Jun 08;18(1):716 [FREE Full text] [doi: 10.1186/s12889-018-5640-1] [Medline: 29884161]
44. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
45. Akhu-Zaheya LM, Shiyab WY. The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases. Int J Med Inform 2017 Feb;98:65-75. [doi: 10.1016/j.ijmedinf.2016.12.003] [Medline: 28034414]
46. Peimani M, Rambod C, Omidvar M, Larijani B, Ghodssi-Ghassemabadi R, Tootee A, et al. Effectiveness of short message service-based intervention (SMS) on self-care in type 2 diabetes: A feasibility study. Prim Care Diabetes 2016 Aug;10(4):251-258. [doi: 10.1016/j.pcd.2015.11.001] [Medline: 26653014]
47. Shahid M, Mahar S, Shaikh S, Shaikh Z. Abstract #224: A Randomized Controlled Trial of Mobile Phone Intervention to Improve Diabetes Care in Rural Areas. Endocrine Pract 2015 May;21(Suppl 2):35. [doi: 10.1016/s1530-891x(20)42530-3]
48. Axley P, Kodali S, Kuo Y, Ravi S, Seay T, Parikh NM, et al. Text messaging approach improves weight loss in patients with nonalcoholic fatty liver disease: A randomized study. Liver Int 2018 May 01;38(5):924-931. [doi: 10.1111/liv.13622] [Medline: 29117472]
49. Barnason S, Zimmerman L, Schulz P, Pullen C, Schuelke S. Weight management telehealth intervention for overweight and obese rural cardiac rehabilitation participants: A randomised trial. J Clin Nurs 2019 May 12;28(9-10):1808-1818 [FREE Full text] [doi: 10.1111/jocn.14784] [Medline: 30667588]
50. Benson GA, Sidebottom A, Hayes J, Miedema MD, Boucher J, Vacquier M, et al. Impact of ENHANCED (diEtitiaNs Helping pAtieNts CarE for Diabetes) telemedicine randomized controlled trial on diabetes optimal care outcomes in patients with type 2 diabetes. J Acad Nutr Diet 2019 Apr;119(4):585-598. [doi: 10.1016/j.jand.2018.11.013] [Medline: 30711463]
51. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
52. Castro Sweet CM, Chiguluri V, Gumpina R, Abbott P, Madero EN, Payne M, et al. Outcomes of a digital health program with human coaching for diabetes risk reduction in a medicare population. J Aging Health 2018 Jun 24;30(5):692-710 [FREE Full text] [doi: 10.1177/0898264316688791] [Medline: 28553807]
53. Choi BG, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Image-based mobile system for dietary management in an American cardiology population: pilot randomized controlled trial to assess the efficacy of dietary coaching delivered via a smartphone app versus traditional counseling. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e10755 [FREE Full text] [doi: 10.2196/10755] [Medline: 31012860]
54. Clark DO, Keith N, Weiner M, Xu H. Outcomes of an RCT of videoconference vs. in-person or in-clinic nutrition and exercise in midlife adults with obesity. Obes Sci Pract 2019 Apr 08;5(2):111-119 [FREE Full text] [doi: 10.1002/osp4.318] [Medline: 31019728]
55. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
56. Haggerty AF, Hagemann A, Barnett M, Thornquist M, Neuhouser ML, Horowitz N, et al. A randomized, controlled, multicenter study of technology-based weight loss interventions among endometrial cancer survivors. Obesity 2017 Nov 31;25(Suppl 2):S102-S108 [FREE Full text] [doi: 10.1002/oby.22021] [Medline: 29086522]
57. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: A randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
58. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and phone-based weight loss intervention: pilot RCT in women at elevated breast cancer risk. Am J Prev Med 2016 Nov;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]

XSL•FO
RenderX

59. Huber JM, Shapiro JS, Wieland ML, Croghan IT, Vickers Douglas KS, Schroeder DR, et al. Telecoaching plus a portion control plate for weight care management: a randomized trial. Trials 2015 Jul 30;16(1):323 [FREE Full text] [doi: 10.1186/s13063-015-0880-1] [Medline: 26223309]
60. Khanna R, Stoddard PJ, Gonzales EN, Villagran-Flores M, Thomson J, Bayard P, et al. An automated telephone nutrition support system for Spanish-speaking patients with diabetes. J Diabetes Sci Technol 2014 Nov 19;8(6):1115-1120 [FREE Full text] [doi: 10.1177/1932296814550186] [Medline: 25239122]
61. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of user- and expert-driven web-based hypertension programs: an RCT. Am J Prev Med 2018 Apr;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
62. Lorig K, Ritter PL, Turner RM, English K, Laurent DD, Greenberg J. A diabetes self-management program: 12-month outcome sustainability from a nonreinforced pragmatic trial. J Med Internet Res 2016 Dec 15;18(12):e322 [FREE Full text] [doi: 10.2196/jmir.6484] [Medline: 27979790]
63. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. J Behav Med 2018 Apr 16;41(2):160-173 [FREE Full text] [doi: 10.1007/s10865-017-9884-5] [Medline: 28918456]
64. McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]
65. McKenzie AL, Hallberg SJ, Creighton BC, Volk BM, Link TM, Abner MK, et al. A novel intervention including individualized nutritional recommendations reduces hemoglobin A1c level, medication use, and weight in type 2 diabetes. JMIR Diabetes 2017 Mar 07;2(1):e5 [FREE Full text] [doi: 10.2196/diabetes.6981] [Medline: 30291062]
66. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
67. Moin T, Damschroder LJ, AuYoung M, Maciejewski ML, Havens K, Ertl K, et al. Results from a trial of an online diabetes prevention program intervention. Am J Prev Med 2018 Nov;55(5):583-591 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.028] [Medline: 30262149]
68. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015 Oct 22;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
69. Nolan RP, Feldman R, Dawes M, Kaczorowski J, Lynn H, Barr SI, et al. Randomized controlled trial of e-counseling for hypertension: REACH. Circ Cardiovasc Qual Outcomes 2018 Jul;11(7):e004420. [doi: 10.1161/CIRCOUTCOMES.117.004420] [Medline: 30006474]
70. Saslow L, Mason AE, Kim S, Goldman V, Ploutz-Snyder R, Bayandorian H, et al. Authors' Reply: Comment on "An online intervention comparing a very low-carbohydrate ketogenic diet and lifestyle recommendations versus a plate method diet in overweight individuals with type 2 diabetes: a randomized controlled trial". J Med Internet Res 2018 May 01;20(5):e181 [FREE Full text] [doi: 10.2196/jmir.8776] [Medline: 29716886]
71. Saslow LR, Summers C, Aikens JE, Unwin DJ. Outcomes of a digitally delivered low-carbohydrate type 2 diabetes self-management program: 1-year results of a single-arm longitudinal study. JMIR Diabetes 2018 Aug 03;3(3):e12 [FREE Full text] [doi: 10.2196/diabetes.9333] [Medline: 30291081]
72. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017 Sep 07;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
73. Vadheim LM, Patch K, Brokaw SM, Carpenedo D, Butcher MK, Helgerson SD, et al. Telehealth delivery of the diabetes prevention program to rural communities. Transl Behav Med 2017 Jun 17;7(2):286-291 [FREE Full text] [doi: 10.1007/s13142-017-0496-y] [Medline: 28417426]
74. Ventura Marra M, Lilly C, Nelson K, Woofter D, Malone J. A pilot randomized controlled trial of a telenutrition weight loss intervention in middle-aged and older men with multiple risk factors for cardiovascular disease. Nutrients 2019 Jan 22;11(2):229 [FREE Full text] [doi: 10.3390/nu11020229] [Medline: 30678197]
75. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
76. Alencar MK, Johnson K, Mullur R, Gray V, Gutierrez E, Korosteleva O. The efficacy of a telemedicine-based weight loss program with video conference health coaching support. J Telemed Telecare 2019 Apr;25(3):151-157. [doi: 10.1177/1357633X17745471] [Medline: 29199544]
77. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]

XSL•FO
RenderX

78. Das SK, Brown C, Urban LE, O'Toole J, Gamache MMG, Weerasekara YK, et al. Weight loss in videoconference and in-person iDiet weight loss programs in worksites and community groups. Obesity 2017 Jun 28;25(6):1033-1041. [doi: 10.1002/oby.21854] [Medline: 28452452]
79. Hageman P, Pullen C, Hertzog M, Pozehl B, Eisenhauer C, Boeckner L. Web-based interventions alone or supplemented with peer-led support or professional email counseling for weight loss and weight maintenance in women from rural communities: results of a clinical trial. J Obes 2017;2017:1602627. [doi: 10.1155/2017/1602627] [Medline: 28480078]
80. Martin CK, Miller AC, Thomas DM, Champagne CM, Han H, Church T. Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial. Obesity 2015 May 28;23(5):935-942. [doi: 10.1002/oby.21063] [Medline: 25919921]
81. Miller KE, Martz DC, Stoner C, Jowers A, Taheri ML, Sarzynski MA, et al. Efficacy of a telephone-based medical nutrition program on blood lipid and lipoprotein metabolism: Results of Our Healthy Heart. Nutr Diet 2018 Feb 07;75(1):73-78. [doi: 10.1111/1747-0080.12366] [Medline: 29411495]
82. Nepper MJ, McAtee JR, Wheeler L, Chai W. Mobile phone text message intervention on diabetes self-care activities, cardiovascular disease risk awareness, and food choices among type 2 diabetes patients. Nutrients 2019 Jun 11;11(6):1314 [FREE Full text] [doi: 10.3390/nu11061314] [Medline: 31212683]
83. Fernandes BSM, Reis IA, Torres HDC. Evaluation of the telephone intervention in the promotion of diabetes self-care: a randomized clinical trial. Rev Lat Am Enfermagem 2016 Aug 29;24:e2719 [FREE Full text] [doi: 10.1590/1518-8345.0632.2719] [Medline: 27579926]
84. Aalbers T, Qin L, Baars MA, de Lange A, Kessels RP, Olde Rikkert MG. Changing behavioral lifestyle risk factors related to cognitive decline in later life using a self-motivated eHealth intervention in Dutch adults. J Med Internet Res 2016 Jun 17;18(6):e171 [FREE Full text] [doi: 10.2196/jmir.5269] [Medline: 27317506]
85. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a nine-month web- and app-based workplace intervention to promote healthy lifestyle and weight loss for employees in the social welfare and health care sector: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e108 [FREE Full text] [doi: 10.2196/jmir.6196] [Medline: 28396303]
86. Bentley CL, Otesile O, Bacigalupo R, Elliott J, Noble H, Hawley MS, et al. Feasibility study of portable technology for weight loss and HbA1c control in type 2 diabetes. BMC Med Inform Decis Mak 2016 Jul 15;16(1):92 [FREE Full text] [doi: 10.1186/s12911-016-0331-2] [Medline: 27418275]
87. Dennison L, Morrison L, Lloyd S, Phillips D, Stuart B, Williams S, et al. Does brief telephone support improve engagement with a web-based weight management intervention? Randomized controlled trial. J Med Internet Res 2014 Mar 28;16(3):e95 [FREE Full text] [doi: 10.2196/jmir.3199] [Medline: 24681761]
88. Gomez-Marcos MA, Patino-Alonso MC, Recio-Rodriguez JI, Agudo-Conde C, Romaguera-Bosch M, Magdalena-Gonzalez O, on behalf the EVIDENT Investigators 11. Short- and long-term effectiveness of a smartphone application for improving measures of adiposity: A randomised clinical trial - EVIDENT II study. Eur J Cardiovasc Nurs 2018 Aug;17(6):552-562. [doi: 10.1177/1474515118761870] [Medline: 29488798]
89. Haas K, Hayoz S, Maurer-Wiesner S. Effectiveness and feasibility of a remote lifestyle intervention by dietitians for overweight and obese adults: pilot study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12289 [FREE Full text] [doi: 10.2196/12289] [Medline: 30973338]
90. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with Type 2 diabetes, Part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
91. Järvelä-Reijonen E, Karhunen L, Sairanen E, Muotka J, Lindroos S, Laitinen J, et al. The effects of acceptance and commitment therapy on eating behavior and diet delivered through face-to-face contact and a mobile app: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Feb 27;15(1):22 [FREE Full text] [doi: 10.1186/s12966-018-0654-8] [Medline: 29482636]
92. Kanera IM, Willems RA, Bolman CAW, Mesters I, Verboon P, Lechner L. Long-term effects of a web-based cancer aftercare intervention on moderate physical activity and vegetable consumption among early cancer survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Feb 10;14(1):19 [FREE Full text] [doi: 10.1186/s12966-017-0474-2] [Medline: 28187725]
93. Kempf K, Altpeter B, Berger J, Reuß O, Fuchs M, Schneider M, et al. Efficacy of the telemedical lifestyle intervention program TeLiPro in advanced stages of type 2 diabetes: a randomized controlled trial. Diabetes Care 2017 Jul 12;40(7):863-871. [doi: 10.2337/dc17-0303] [Medline: 28500214]
94. Luley C, Blaik A, Götz A, Kicherer F, Kropf S, Isermann B, et al. Weight loss by telemonitoring of nutrition and physical activity in patients with metabolic syndrome for 1 year. J Am Coll Nutr 2014 Aug 08;33(5):363-374. [doi: 10.1080/07315724.2013.875437] [Medline: 25105874]
95. Orlandoni P, Jukic Peladic N, Spazzafumo L, Venturini C, Cola C, Sparvoli D, et al. Utility of video consultation to improve the outcomes of home enteral nutrition in a population of frail older patients. Geriatr Gerontol Int 2016 Jun 14;16(6):762-767. [doi: 10.1111/ggi.12551] [Medline: 26171754]

XSL•FO
RenderX

96. Recio-Rodriguez J, Agudo Conde C, Calvo-Aponte M, Gonzalez-Viejo N, Fernandez-Alonso C, Mendizabal-Gallastegui N, EVIDENT Investigators. The effectiveness of a smartphone application on modifying the intakes of macro and micronutrients in primary care: a randomized controlled trial. The EVIDENT II Study. Nutrients 2018 Oct 10;10(10):1473 [FREE Full text] [doi: 10.3390/nu10101473] [Medline: 30309008]
97. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MDC, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
98. van Doorn-van Atten MN, Haveman-Nies A, van Bakel MM, Ferry M, Franco M, de Groot LCPGM, et al. Effects of a multi-component nutritional telemonitoring intervention on nutritional status, diet quality, physical functioning and quality of life of community-dwelling older adults. Br J Nutr 2018 May 15;119(10):1185-1194. [doi: 10.1017/s0007114518000843]
99. Elbert SP, Dijkstra A, Oenema A. A mobile phone app intervention targeting fruit and vegetable consumption: the efficacy of textual and auditory tailored health information tested in a randomized controlled trial. J Med Internet Res 2016 Jun 10;18(6):e147 [FREE Full text] [doi: 10.2196/jmir.5056] [Medline: 27287823]
100. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and HbA1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (The ANODE Study). J Med Internet Res 2017 Nov 08;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
101. Whitelock V, Kersbergen I, Higgs S, Aveyard P, Halford JCG, Robinson E. A smartphone based attentive eating intervention for energy intake and weight loss: results from a randomised controlled trial. BMC Public Health 2019 May 21;19(1):611 [FREE Full text] [doi: 10.1186/s12889-019-6923-x] [Medline: 31113400]
102. Fletcher J. What is heterogeneity and is it important? BMJ 2007 Jan 13;334(7584):94-96 [FREE Full text] [doi: 10.1136/bmj.39057.406644.68] [Medline: 17218716]
103. Sweet SN, Fortier MS. Improving physical activity and dietary behaviours with single or multiple health behaviour interventions? A synthesis of meta-analyses and reviews. Int J Environ Res Public Health 2010 Apr 16;7(4):1720-1743 [FREE Full text] [doi: 10.3390/ijerph7041720] [Medline: 20617056]
104. Helf C, Hlavacs H. Apps for life change: Critical review and solution directions. Entertain Comput 2016 May;14:17-22. [doi: 10.1016/j.entcom.2015.07.001]
105. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]

Abbreviations

CVD: cardiovascular disease
HbA_{1c}: glycosylated hemoglobin (A_{1c})
MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA-P 2015: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015
RCT: randomized controlled trial

Edited by G Eysenbach, R Kukafka; submitted 25.07.19; peer-reviewed by C Yuan, M Bardus, L Saslow, M Karunanithi; comments to author 04.11.19; revised version received 28.03.20; accepted 12.04.21; published 17.05.21.

Please cite as:
Robert C, Erdt M, Lee J, Cao Y, Naharudin NB, Theng YL
Effectiveness of eHealth Nutritional Interventions for Middle-Aged and Older Adults: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(5):e15649
URL: https://www.jmir.org/2021/5/e15649
doi:10.2196/15649
PMID:33999005

©Caroline Robert, Mojisola Erdt, James Lee, Yuanyuan Cao, Nurhazimah Binte Naharudin, Yin-Leng Theng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

XSL•FO
RenderX

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

Internet-Based Cognitive Behavioral Therapy for Patients Reporting Symptoms of Anxiety and Depression After Myocardial Infarction: U-CARE Heart Randomized Controlled Trial Twelve-Month Follow-up

Sophia Monica Humphries[1], MSc; John Wallert[1,2,3], PhD; Fredrika Norlund[1], PhD; Emma Wallin[1,4], PhD; Gunilla Burell[1,5], PhD; Louise von Essen[1], PhD; Claes Held[1,6], MD, PhD; Erik Martin Gustaf Olsson[1], PhD

[1]Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

[3]Centre for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

[4]Department of Psychology, Uppsala University, Uppsala, Sweden

[5]Family Medicine and Preventative Medicine, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[6]Department of Medical Sciences, Cardiology, Uppsala Clinical Research Centre, Uppsala University, Uppsala, Sweden

Corresponding Author:
Sophia Monica Humphries, MSc
Clinical Psychology in Healthcare
Department of Women's and Children's Health
Uppsala University
Akademiska Hospital
Uppsala, 75185
Sweden
Phone: 46 760341186
Email: sophia.humphries@kbh.uu.se

Abstract

Background: The U-CARE Heart trial was one of the first randomized controlled trials to evaluate the effect of internet-based cognitive behavioral therapy on self-reported symptoms of anxiety or depression for patients with a recent myocardial infarction. While the effects of internet-based cognitive behavioral therapy on Hospital Anxiety and Depression Scale (HADS) scores at 14 weeks postbaseline were not significant, in this study, we investigated possible long-term effects of treatment.

Objective: The aim of this study was to evaluate the long-term effectiveness of internet-based cognitive behavioral therapy on self-reported symptoms of anxiety and depression in patients 12 months after a myocardial infarction and to explore subsequent occurrences of cardiovascular disease events.

Methods: Shortly after acute myocardial infarction, 239 patients (33% female, mean age 59.6 years) reporting mild-to-moderate symptoms of anxiety or depression were randomized to 14 weeks of therapist-guided internet-based cognitive behavioral therapy (n=117) or treatment as usual (n=122). Data from national registries were used to explore group differences in clinical outcomes such as cardiovascular disease and cardiovascular-related mortality for a follow-up period of up to 5 years: group differences in HADS total score 1 year post–myocardial infarction, the primary outcome, was analyzed using multiple linear regression. Secondary outcomes, such as HADS anxiety and depression subscales and the Cardiac Anxiety Questionnaire total score (CAQ), which measures heart-focused anxiety, were analyzed in the same way. Multiple imputation was used to account for missing data, and a pooled treatment effect was estimated. Adjusted Cox proportional hazards models were used to estimate hazard ratios (HRs) for data pertaining to registry outcomes.

Results: Both groups reported lower HADS total scores 1 year after myocardial infarction than those at baseline. HADS total scores were not significantly different between the treatment and control groups 1 year after myocardial infarction (β=–1.14, 95% CI –2.73 to 0.45, *P*=.16). CAQ was the only measure improved significantly by internet-based cognitive behavioral therapy when compared with treatment as usual (β=–2.58, 95% CI –4.75 to –0.42, *P*=.02) before adjusting for multiple comparisons. The composite outcome of nonfatal cardiovascular events and cardiovascular-related mortality did not differ between groups but was numerically higher in the internet-based cognitive behavioral therapy group, who were at slightly greater risk (HR 1.8, 95% CI

XSL•FO
RenderX

0.96 to 3.4, *P*=.07). Adjusting for previous myocardial infarction and diabetes attenuated this estimate (HR 1.5, 95% CI 0.8 to 2.8, *P*=.25).

Conclusions: Internet-based cognitive behavioral therapy was not superior in reducing self-reported symptoms of depression or anxiety compared to treatment as usual at the 1-year follow-up after myocardial infarction. A reduction in cardiac-related anxiety was observed but was not significant after adjusting for multiple comparisons. There was no difference in risk of cardiovascular events between the treatment groups. Low treatment adherence, which might have affected treatment engagement and outcomes, should be considered when interpreting these results.

Trial Registration: ClinicalTrials.gov NCT01504191; https://clinicaltrials.gov/ct2/show/NCT01504191

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-015-0689-y

(J Med Internet Res 2021;23(5):e25465) doi:10.2196/25465

KEYWORDS

myocardial infarction; iCBT; psychological treatment; cardiovascular health; cognitive behavior therapy; internet; cardiovascular; infarction; treatment; anxiety; depression

Introduction

Myocardial infarction continues to be the leading cause of death worldwide [1]. The risk increases with age and is associated with behavioral factors such as smoking and reduced physical activity [2]. For over one-third of patients after myocardial infarction, symptoms of anxiety, depression, or both remain elevated 3 months after hospital discharge [3], and similar levels have even been reported up to 1 year after myocardial infarction [4]. A preexisting history of mental illness, such as depression, is associated with a 31% increased risk of a recurrent myocardial infarction [5] and a 30% higher mortality rate in the 18 months following myocardial infarction than among those without a history of depression [6]. Mortality rates increase with severity of depression, according to one study [7] and, depressive symptoms are reported to be more prevalent among patients with more cardiovascular risk factors (eg, smoking and obesity) [8]. This combination of increased risk for morbidity following a myocardial infarction and lower reported quality of life [9,10] makes therapy that is aimed at reducing symptoms of anxiety and depression likely a beneficial treatment option. In a recent meta-analysis [11], psychological treatment for patients with coronary heart disease was linked to a reduction in self-reported symptoms of anxiety and depression and reduced risk of cardiac-related mortality compared to treatment as usual.

Web-based interventions offer remote access to treatment which are sometimes otherwise inaccessible [12] and have been evaluated against face-to-face therapy in randomized controlled trials with varying yet promising results [12-14]. Some trials have used therapy conducted online, often in the form of internet-based cognitive behavioral therapy, to treat depression in patients with heart failure [15], cardiovascular disease [16], and with high-cardiovascular disease risk [17], but before the U-CARE Heart trial [18], no study had offered internet-based cognitive behavioral therapy specifically to patients after myocardial infarction. A more recent study [19] of internet-based cognitive behavioral therapy for symptoms of anxiety and depression in patients who had experienced an acute coronary event, including myocardial infarction, demonstrated a strong effect of treatment on reduction of these symptoms that was sustained at a 4-week follow-up.

The U-CARE heart trial [18] was, to the authors' knowledge, the first randomized controlled trial to offer internet-based cognitive behavioral therapy to patients experiencing mild to moderate self-reported symptoms of anxiety or depression after a myocardial infarction. Posttreatment follow-up at 14 weeks found an overall reduction in symptoms of anxiety and depression, according to the Hospital Anxiety and Depression Scale (HADS) [20] compared to those at baseline, yet no differences were found between the intervention and control group. Long-term treatment follow-up of the study population was intended to gather more information about the sustainability of general improvements in anxiety and depression and whether group characteristics (symptoms of anxiety and depression, myocardial infarction and stroke events, and cardiovascular-related mortality) differed at follow-up. The ultimate goal of a psychological intervention post–myocardial infarction would be, not only to reduce symptoms of anxiety and depression long-term but also, to reduce cardiovascular morbidity and mortality, although this was not the main purpose of the U-CARE Heart trial.

The purpose of this follow-up study was thus to investigate whether patients with mild-to-moderate symptoms of anxiety or depression, self-reported within 2 months following myocardial infarction and treated with internet-based cognitive behavioral therapy, had (1) improvements in self-reported mental health at 12 months after the myocardial infarction, and (2) reduced risk for cardiovascular events including cardiovascular-related mortality, for a period of up to 5 years, compared to those receiving treatment as usual.

Methods

Study Design and Participants

This follow-up study (1) evaluated the effect of internet-based cognitive behavioral therapy on self-reported symptoms of anxiety and depression 12 months after the myocardial infarction, and (2) explored whether internet-based cognitive behavioral therapy had an effect on nonfatal cardiovascular events or cardiovascular-related mortality up to 5 years after study inclusion. The study design, procedure, intervention, and results of the posttreatment follow-up of the U-CARE Heart trial have been reported [18,21].

XSL•FO
RenderX

In brief, the U-CARE Heart trial was a multicenter study that recruited participants <75 years of age, reporting symptoms of anxiety or depression (scoring >7 on any of the 2 HADS anxiety and depression subscales) within 3 months following their myocardial infarction. Myocardial infarction was defined according to International Statistical Classification of Disease Tenth Revision (ICD-10) code I21 and diagnosed by a cardiologist. Patients scheduled for coronary artery bypass surgery, with low adherence (such as missing appointments with the cardiac nurse or substance use), or expected to live <1 year, as judged on-site by the recruiting nurse, were not eligible. Patients who met eligibility criteria (n=239) were randomized to receive either internet-based cognitive behavioral therapy (n=117) or treatment as usual (n=122). This study received ethical approval by the regional ethics committee in Uppsala (2011/217) and is registered on ClinicalTrials.gov (NCT01504191).

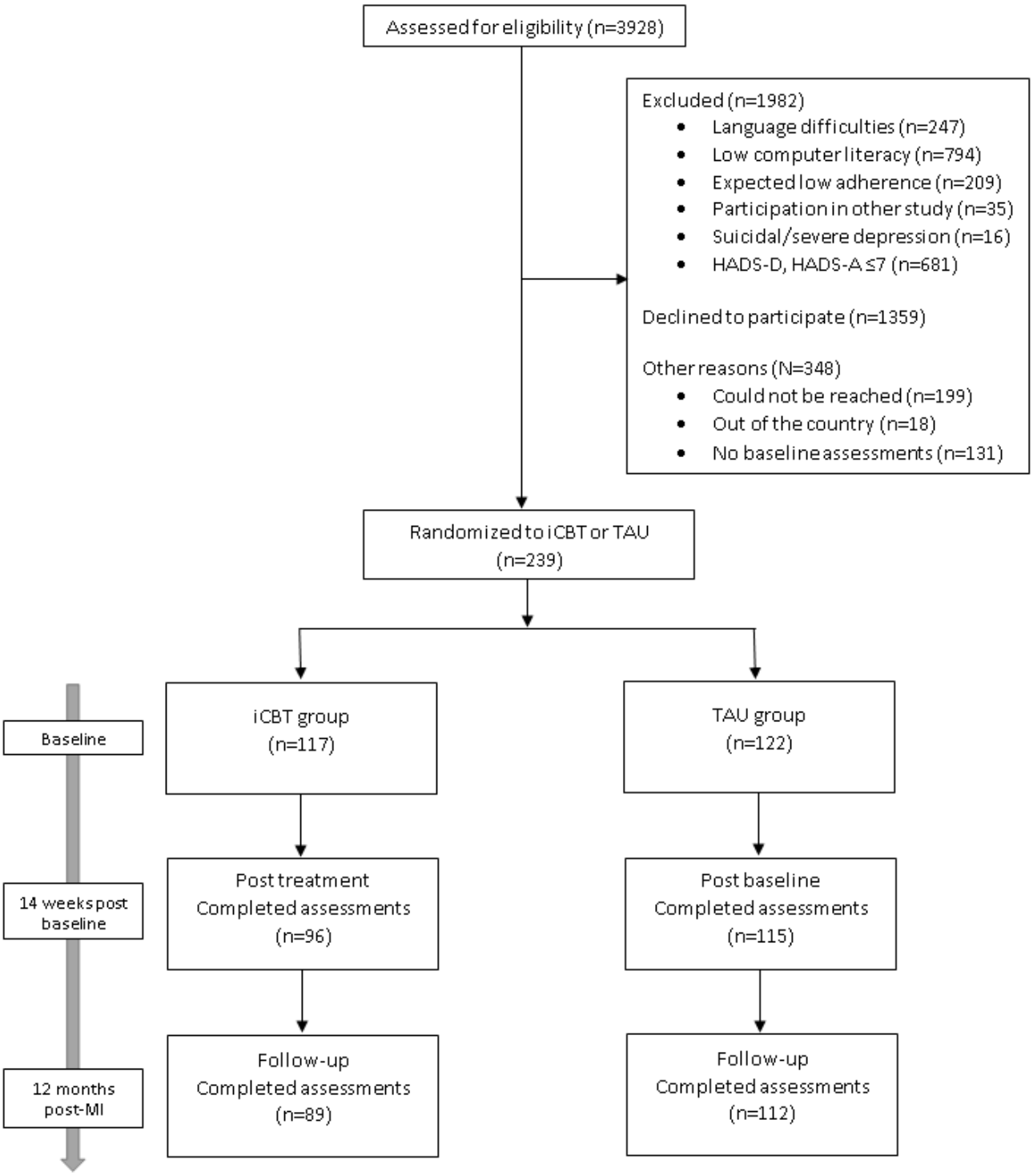
Procedure

Screening for eligible participants took place across in Sweden from September 2013 to December 2016 in 25 cardiac clinics during routine visits between 1 and 8 weeks following their myocardial infarction. Consenting participants were sent an email and password login to a secure internet-based portal (Multimedia Appendix 1) where they completed baseline assessments and would subsequently receive the internet-based cognitive behavioral therapy (if applicable). Eligibility (Figure 1) was assessed in the baseline questionnaires and those scoring >7 on any or both of the HADS subscales were randomized automatically in the internet-based portal 1:1 to internet-based cognitive behavioral therapy or treatment as usual. The full recruitment procedure has previously been reported [18].

Trial-specific outcomes were measured 14 weeks postbaseline, which corresponded to treatment completion for those in the internet-based cognitive behavioral therapy group, and again at follow-up 12 months after myocardial infarction. Participants in both the intervention and treatment-as-usual groups were sent automatic SMS texts and email reminders asking them to fill in the web-based questionnaires at each observation point and were subsequently reminded via telephone by one of the research staff (blind to treatment allocation) if initially unresponsive. Figure 1 details the patient flow through the study.

XSL•FO
RenderX

Figure 1. Participation flowchart. iCBT: internet-based cognitive behavioral therapy; TAU: treatment as usual; HADS-D: Hospital Anxiety and Depression Scale–Depression subscale; HADS-A: Hospital Anxiety and Depression Scale–Anxiety subscale; MI: myocardial infarction.



Intervention

The internet-based cognitive behavioral therapy intervention [18,21] delivered via the U-CARE portal (a secure web-portal) was therapist guided and partly customizable. This meant that after a compulsory introduction module, patients could choose their treatment modules. The intervention was developed for cardiac patients and consisted of 11 modules: Introduction, Managing Worry, Fear and Avoidance, Behavioral Activation, Problem Solving, Communication Skills, Applied Relaxation Training, Managing Negative Thoughts, Coping With Insomnia, Values in Life, and Relapse Prevention. Homework assignments were included in each module, and the portal also included a library that was accessible at all times with content and material including media such as videos and informational text.

All participants received standard protocol secondary prevention and cardiac rehabilitation offered by the regional health care system. This usually includes, but is not limited to, preventive medications, education about cardiovascular risk-factors, smoking cessation, organized and tailored physical exercise, and in some areas, access to counseling or psychosocial support.

XSL•FO
RenderX

Measures

Sociodemographic and Psychological Measures

Sociodemographic data were self-reported at baseline. The primary outcome measure was HADS total score, measuring self-reported symptoms of both anxiety and depression (subscales). The HADS is a useful tool in the screening of depression in cardiac patients [22,23], and there is support of its validity for online use [24].The 2 subscales are divided equally between 14 questions, with 7 questions pertaining to each subscale. Questions are answered on a 4-point scale (0-3), with a score >7 on either subscale indicating at least mild symptoms [20].

Secondary outcome measures included the Cardiac Anxiety Questionnaire (CAQ) [25], which assesses self-reported heart-related anxiety; Montgomery-Åsberg Depression Rating Scale–Self-rated (MADRS-S) [26], which measures depression; and the Behavioral Activation for Depression Scale Short–Form (BADS-SF), which assesses depression-related behaviors [27]. In contrast to the other scales, for which a high score denotes more severe symptoms, a high score on the BADS-SF is favorable as it indicates higher activation of change to depression-related behaviors.

Register Data

Data on cardiovascular events and mortality during follow-up were obtained from the Patient Registry and the Causes of Death Registry of the National Board of Health and Welfare in Sweden [28]. ICD-10 diagnostic codes were used [29]. A composite cardiovascular event outcome was created comprising hospitalization for any of the following diagnoses between time of randomization and point of data censoring: cardiovascular death (ICD cause of death code I), acute coronary syndrome (ICD codes I20, I21, and I22), heart failure (ICD codes I50 and I11.0), stroke (ICD codes I61, I62, I63, and I64), and revascularization (intervention codes FNG00, FNG02, FNG05, FNG10, and FNC). Background data for medical history were taken from the Swedish Web-System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) registers [30]: RIKS-HIA (the National Register for Information and Knowledge on Heart Intensive Care Admission) and SEPHIA (the Registry for Secondary Prevention after Heart Intensive Care Admission). These sources of data are linked via a personal identification number that is unique to every registered resident in Sweden, matched to the study cohort with codes, and handled in accordance with current data protection standards.

Statistical Analysis

Statistical analysis was planned in line with CONSORT (Consolidated Standards of Reporting Trial) before treatment allocation was disclosed and before the trial database was locked.

Main analysis investigated potential differences in HADS total scores 12 months post–myocardial infarction, with subsequent analyses of the HADS anxiety and HADS depression subscale scores. Secondary analyses focused on potential differences between groups in MADRS-S total score, CAQ total score, and BADS-SF total score.

Linear modeling was used to estimate treatment effects. As with the posttreatment analysis [18], analyses were conducted based on intention to treat. Treatment was an independent variable and 12-month follow-up HADS total score as the dependent variable, while controlling for baseline HADS total score, sex, and age. This adjustment for covariates was applied to all analyses. The HADS subscales were analyzed in the same way but only with patients included in the study based on the specific subscale. Thus, it was possible for patients to be included in both analyses for both subscales if they scored >7 on both at baseline.

To account for missing values and analyze according to intention to treat, data were imputed by multiple imputation using chained equations and predictive mean matching [31]. Variables used in the imputation model were the same as those in the posttreatment analysis [18], which were chosen on the basis of potential association with the main outcome and with input from cardiologist and psychologist recommendations. Unless specified, reported results are based on imputed data.

HADS total score sensitivity analyses were conducted on observed data (ie, nonimputed). Supplementary analyses of HADS total score were based on per protocol data, including only those in the model who had completed at least one homework assignment. Only the analyses according to intention to treat (with imputed data) were applied to the secondary outcomes. Effect estimates are reported as pooled adjusted point estimates β with 95% confidence intervals. Paired *t* tests were conducted with outcomes pre- vs posttreatment to assess change over time across the full sample. Level of significance, α, for all separate comparisons was set to $P<.05$. For familywise testing of the multiple hypotheses, a Bonferroni-corrected *P* value=.008 (6 hypotheses) was used.

Multivariable Cox regression analysis was performed to estimate adjusted hazard ratios (HR) with 95% confidence intervals with time (in days) to first cardiovascular event as the outcome event. The exposure was treatment group, with time under risk starting from the date of randomization during follow-up and censoring at the time of any cardiovascular event, end of follow-up (December 31, 2018), or the date of noncardiovascular death, whichever came first.

The first participant was randomized in September 2013. The longest period of follow-up spanned 1898 days, and the shortest spanned 719 days; the mean follow-up period was 1139.3 days. Any incidence of cardiovascular death or cardiovascular disease event was used as the primary outcome for time-to-event analysis. There were no missing data for these outcomes. Adjustments for age and sex were performed for higher precision. Posthoc exploratory analyses were conducted with previous myocardial infarction and diabetes as additional covariates since previous medical history of these appeared to be more prevalent in the treatment group. All statistical analyses were performed using R statistical software (version 4.0.2, The R Project).

XSL•FO
RenderX

Missing Data

Of 239 patients who were enrolled, 24% (28 participants) from the internet-based cognitive behavioral therapy group and 8% (10 participants) from the treatment-as-usual group had missing data on the HADS questionnaire at the 12-month follow-up. In the internet-based cognitive behavioral therapy and treatment-as-usual groups 26% (31 participants) and 9% (11 participants), respectively, had missing data on MADRS-S; 26% (30 participants) and 8% (10 participants), respectively, had missing CAQ values; and 30% (35 participants) and 11% (13 participants) respectively, had missing data on the BADS-SF.

Data were considered missing if (1) the entire questionnaire was left unanswered (ie, the participant skipped the questionnaire or the participant did not log in or respond to any of the questionnaires at follow-up) or (2) items were skipped within a questionnaire that were required in order to compute a value for the total score (eg, missing items within the HADS questionnaire, which made it impossible to compute a total score from the observed data).

Results

Patient Characteristics

Baseline characteristics are presented in Table 1. Overall, patients were a mean age of 59.6 years (SD 8.49), and 33.5% were women (80/239). At baseline, 24.3% (58/239) reported to be receiving other forms of counseling and 18.0% (39/239) of the study sample reported taking psychotropic medications. At baseline, there was no difference between the groups in the mean HADS total score (P=.57).

All primary and secondary measures (Figure 2, Multimedia Appendix 2) at posttreatment and follow-up displayed changes from baseline, indicating a reduction of self-reported anxiety or depression symptoms.

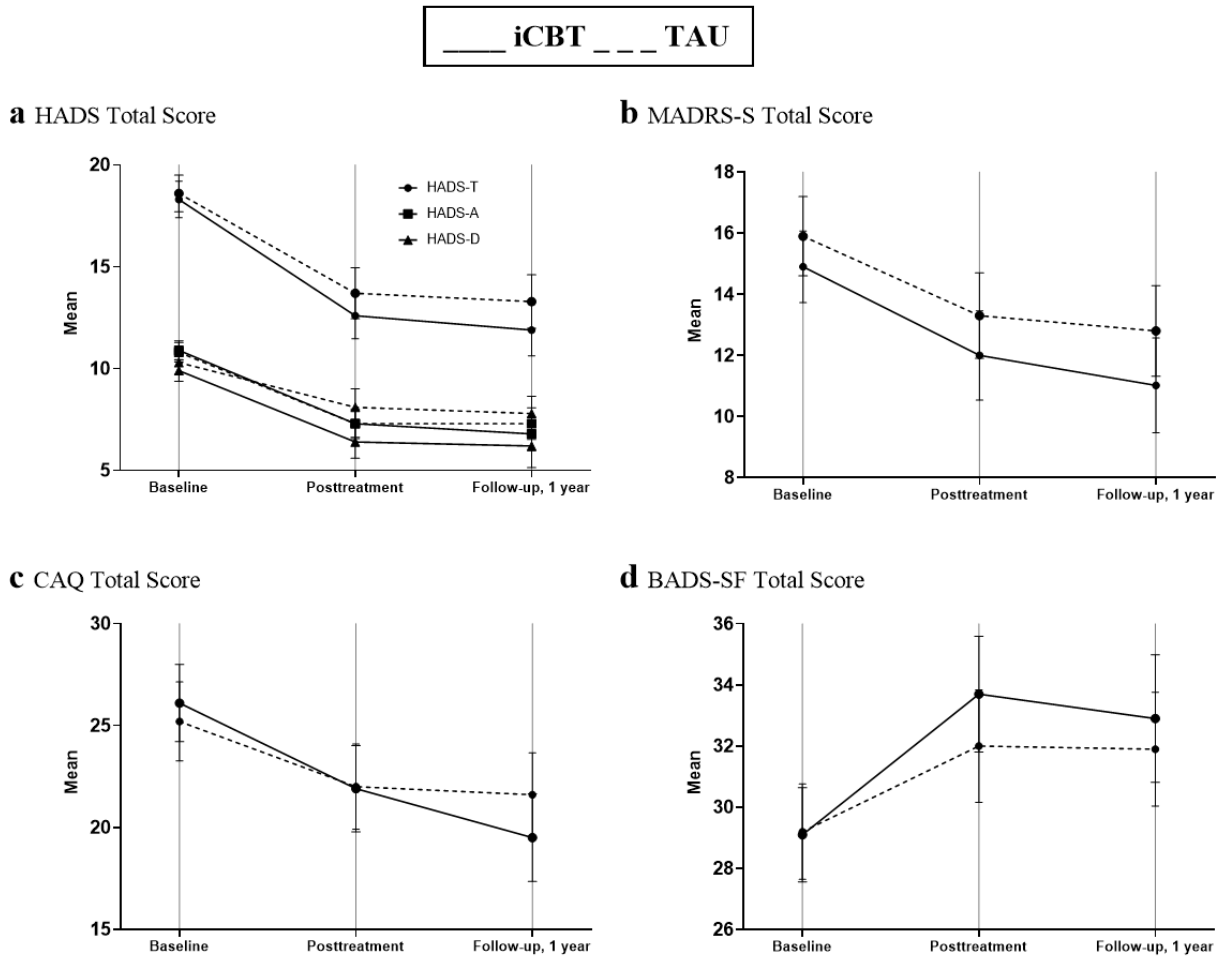
XSL•FO
RenderX

Table 1. Patient characteristics with observed data (ie, nonimputed).

Sociodemographic characteristics	Internet-based cognitive behavioral therapy (n=117)	Treatment as usual (n=122)
Age (years), mean (SD)	58.4 (9.0)	60.8 (7.8)
Gender, n (%)		
Women	44 (37.6)	36 (29.5)
Men	73 (62.4)	86 (70.5)
Occupation, n (%)		
Employed	78 (66.7)	66 (54.1)
Unemployed	4 (3.4)	2 (1.6)
Retired	33 (28.2)	37 (30.3)
Sick leave	2 (1.7)	1 (0.8)
Other	0 (0.0)	2 (1.6)
Highest attained education, n (%)		
Primary school	22 (18.8)	26 (21.3)
Secondary school	45 (38.5)	46 (37.7)
University	50 (42.7)	50 (41.0)
In a relationship, n (%)		
Yes	99 (84.6)	101 (82.8)
No	18 (15.4)	21 (17.2)
Smoker, n (%)	6 (5.1)	8 (6.6)
Yes		
No		
Body mass index (kg/m^2), mean (SD)	27.8 (5.0)	27.4 (4.0)
Taking psychotropic medicine, n (%)		
Yes	19 (16.2)	20 (16.4)
No	98 (83.8)	102 (83.6)
Other current counseling, n (%)		
Yes	30 (25.7)	28 (22.9)
No	87 (74.3)	94 (77.1)
Medical history, n (%)		
Myocardial infarction	19 (16.2)	13 (10.7)
Diabetes	21 (17.9)	19 (15.6)
Hypertension	42 (35.9)	51 (41.8)
Hyperlipidemia	26 (22.2)	27 (22.1)
Stroke	0 (0.0)	4 (3.3)
Heart failure	4 (3.4)	2 (1.6)

XSL•FO
RenderX

Figure 2. (a) Hospital Anxiety and Depression Scale total score, (b) Montgomery-Åsberg Depression Scale–Self-rated total score, (c) Cardiac Anxiety Questionnaire total score, and (d) Behavioral Activation for Depression Scale–Short Form total score mean (observed data) values at baseline, posttreatment, and 1-year follow-up. Error bars indicate 95% CI. iCBT: internet-based cognitive behavioral therapy; TAU: treatment as usual.



Main Analysis

HADS total scores from baseline to the 12-month follow-up visit after myocardial infarction showed a consistent reduction across the total study sample (mean difference 5.7; t_{200}=13.57, *P*<.001). Figure 2 shows changes in mean scores over the 3 time points (baseline, posttreatment, and 12-month follow-up) of the observed (nonimputed) data.

The main model showed no difference of internet-based cognitive behavioral therapy on HADS total scores at follow-up (β=–1.14, 95% CI –2.73 to 0.45, *P*=.16). There was no effect of internet-based cognitive behavioral therapy on HADS depression (β=–1.08, 95% CI –2.31 to 0.16, *P*=.09) or anxiety (β=–0.64, 95% CI –1.65 to 0.37, *P*=.21) subscale scores.

Secondary Outcomes

The CAQ was the only additional outcome to show a significant reduction at the 12-month follow-up (β=–2.58, 95% CI –4.75 to –0.42, *P*=.02) for internet-based cognitive behavioral therapy; however, this was not significant after adjusting for multiple comparisons (Bonferroni-corrected level of significance *P*<.008). Further details on outcomes are available in Multimedia Appendix 2.

Sensitivity analysis of observed data (β=–1.3, 95% CI –2.96 to 0.27, *P*=.10) and per protocol analysis (β=–1.3, 95% CI –2.90 to 0.41, *P*=.14) showed no effect of internet-based cognitive behavioral therapy on HADS total score.

Cardiovascular Clinical Outcomes

Compared to those in the treatment-as-usual group, participants receiving internet-based cognitive behavioral therapy had a numerically but not significantly higher risk of experiencing a cardiovascular event (HR 1.8, 95% CI 0.96 to 3.4, *P*=.07). Posthoc analysis showed this estimate was attenuated when previous myocardial infarction and diabetes, as indicators of baseline health status, were entered in the model (HR 1.5, 95% CI 0.8 to 2.8, *P*=.25). Cardiovascular events included in the analysis are shown in Table 2.

XSL•FO
RenderX

Table 2. Number of patients experiencing cardiovascular events before censoring (first event), including cardiovascular-related death during follow-up.

ICD[a] classification	Patients, n (first event, n)		
	All	Internet-based cognitive behavioral therapy	Treatment as usual
Total	61	36 (25)	25 (17)
Cardiovascular-related death	2	1 (0)	1 (0)
Acute coronary syndrome	23	14 (13)	9 (8)
Stroke	3	2 (2)	1 (1)
Revascularization	24	12 (4)	12 (7)
Heart failure	9	7 (6)	2 (1)

[a]ICD: International Statistical Classification of Diseases.

Discussion

Principal Findings

Primary analysis did not detect a significant difference in HADS total score between the treatment and control groups at follow-up 12 months after the myocardial infarction. These findings were in line with those of the original paper (presenting posttreatment results, ie, 14 weeks after baseline [18]). Of the secondary measures, the CAQ was the only outcome to show an effect of treatment in mean scores between the internet-based cognitive behavioral therapy and treatment-as-usual group at the 12-month follow-up, and occurrences of cardiovascular-related events did not differ significantly between the groups, although statistical power was limited.

Possible explanations to why there was no difference observed between the groups are low activity-levels among patients in the internet-based cognitive behavioral therapy arm and the low number of completed treatment modules. Just over half of those in the internet-based cognitive behavioral therapy group completed the first introductory module, and only 15% continued to work through any of the remaining 10 modules [18]. Interviews with patients from this study in a follow-up study [32] identified several factors that potentially contributed to low treatment adherence: lack of time, technical aspects (eg, insufficient computer literacy), and unpleasant emotions evoked by the intervention.

The intervention design may have contributed to low adherence. Creating an internet-based cognitive behavioral therapy intervention is technically challenging and complex and must be made with the characteristics and demographics, such as age, of the population in mind. Despite excluding individuals not able or willing to use a computer, the internet, email, or a mobile device, participants still reported low computer literacy as a contributing factor to why they did not engage more with the intervention [32]. A study [16] that reported high adherence to internet-based cognitive behavioral therapy for patients with cardiovascular disease (60% completed all modules) used nurses with experience of cardiovascular disease care to guide the intervention and suggested that this was beneficial as it allowed the participants to ask cardiovascular disease-related questions and health-related concerns. Providing a similar option for patients in our study could have likewise improved adherence. Moreover, age is one of the most well-known risk factors for myocardial infarction [33] so it is not unexpected that patients are older on average than other patient populations treated with internet-based cognitive behavioral therapy; it is important to understand how demographic characteristics of myocardial infarction patients might affect treatment engagement, particularly as it has been found that one-third of adults aged over 65 years who were average or modest internet users report feeling anxious around technology [34]. Therefore, studies with low treatment attrition and high participant satisfaction that have tailored certain aspects towards populations that may be less familiar with digital technology, should be used as design examples. For example, a recent pilot study [35] of treatment for depression in healthy older adults provided tablet-device training to the participants beforehand, required no internet connection, and used a stylus to operate the touch screen. Applying these elements to the intervention may improve results.

A well-discussed issue in intervention literature that applies to this study concerns how participants are recruited into a study that offers treatment. Patients were approached by hospital staff in participating hospitals and asked if they would be interested in participating rather than responding to an open-advertisement or actively seeking care themselves. Similar studies have also suggested that clinical recruitment might have contributed to low internet-based cognitive behavioral therapy adherence [36]. It is likely that self-referring patients would be more motivated to complete psychological treatments. A meta-analysis [37] found a more favorable effect of internet-based cognitive behavioral therapy (vs waitlist control) on anxiety outcomes for studies that deployed open community recruitment strategies as opposed to clinical service recruitment. One could speculate that this favorable difference was due to greater treatment adherence of self-referring patients and stricter inclusion and exclusion criteria.

More recent trials have had better success in maintaining participant retention, particularly when patients could self-identify as needing help with symptoms related to their coronary event and when there were fewer treatment modules to work through [19].

Of note, secondary analyses revealed significant long-term improvement on the CAQ scale in the internet-based cognitive behavioral therapy group compared to that of the treatment-as-usual group (P=.02). Interestingly, this parameter did not differ between groups at 14 weeks postbaseline [18].

XSL•FO
RenderX

One possible explanation might be that treatment was tailored relatively well to cardiac-specific scenarios and provided examples that were well-suited to someone with a recent myocardial infarction. Machine learning modeling with the U-CARE Heart trial data has since shown that high CAQ total score predicts participants' adherence to the treatment [38] and that the CAQ fear subscale was the strongest predictor of participant adherence. The group differences observed on these variables might explain the intrinsic motivation of some participants to adhere more than others, and thus, may explain the findings in the CAQ outcome. Since so few participants adhered to the treatment, however, we cannot conclude that this was the reason for better CAQ scores. Previous research [4,39] has demonstrated that a treatment effect might appear at a later follow-up period despite that no effect being observed immediately posttreatment. This may explain why these effects were not present 14 weeks postbaseline. Multiple hypothesis testing can also become a problem as some results may appear by chance. After adjusting multiple comparisons, the CAQ *P* value (*P*=.02) was not significant.

The risk of cardiovascular events did not differ across groups at long-term follow-up. Unexpectedly, the hazard ratio indicated a slightly negative effect of the treatment. This effect was not significant (*P*=.07); this was a small study, with few events and lacking power for this exploratory analysis. Despite this, such an analysis is arguably justified as the knowledge gained thereof, and its potential benefits could contribute to the literature of internet-based cognitive behavioral therapy effects on cardiovascular events. Since we included a rather broad range of cardiovascular events, it is possible that those in the internet-based cognitive behavioral therapy group experienced more but milder incidences such as heart failure; the number of events were too few to adequately explore this, but upon counting the different event types in the 2 groups, no particular pattern emerged. It could also be that the treatment group was less healthy at baseline as randomization does not necessarily make groups similar. When adding indications of baseline health differences (previous myocardial infarction and diabetes), the effect was attenuated. A similar recent study [40] also found no difference in clinical outcomes between a control group and a cognitive behavioral therapy/well-being therapy group. In our study, one participant in each group died, but since neither death was the first event experienced by each respective participant, these deaths were excluded from analysis. However, as noted above, since adherence was low, it is difficult to conclude whether any effects were due to the treatment. One could speculate that it is distressing for some people to participate in a treatment study where activity is expected as it can engage feelings of guilt due to low adherence. Increased distress was the main interpretation of the negative effect on mortality in women in the nurse-led intervention in the M-HART trial [41].

Limitations

The inclusion of participants with a rather low level of self-reported anxiety or depression (HADS anxiety or HADS depression subscale scores >7) meant that patients with only mild symptoms were included, which could be considered a disadvantage—this may have contributed to the lack of difference in the primary outcome both immediately posttreatment and at the 1-year follow-up. Spontaneous improvement over time in anxiety and depression has been observed in patients with acute myocardial infarction [42,43]. At 2 months post–myocardial infarction, it may have been too early to know which patients would still report problematic symptom levels versus those whom would recover over time, particularly as some high anxiety–reporting myocardial infarction patients have been found to steadily improve between 3 and 6 months post–myocardial infarction [44]. As previously reported [21], problems with recruitment were the initial reason for lowering the HADS anxiety and HADS depression inclusion scores from >10 to >7, and more stringent exclusion criteria (ie, a higher HADS anxiety or HADS depression threshold value for inclusion) may have allowed stronger treatment effects to be observed.

As previously reported [18], the HADS was developed largely as a screening measure for anxiety and depression and might not have been sensitive enough to detect minor changes across the measurement time points, although the inclusion of multiple secondary measures may have somewhat made up for this. Using a diagnostic interview technique is arguably one of the most robust methods to screen for anxiety and depression; however, this would have required a lot more time and therapist involvement, been more expensive, and been further from the digitalization goals around which this treatment was structured.

One challenge in this study, as is often the case in long-term follow-up studies, was the amount of missing data due to nonresponses. There was also more attrition in the internet-based cognitive behavioral therapy group than in the treatment-as-usual group at the 1-year follow-up. While this would need to be explored further to fully understand the reasons for the differences between the groups, one could speculate that it was partly because of the perceived burden by those in the internet-based cognitive behavioral therapy group of completing so many additional web-based forms and exercises, which was not experienced by those in the treatment-as-usual group. Alternatively, patients assigned to the internet-based cognitive behavioral therapy intervention could have been randomly more ill from the start, a factor which could contribute to more attrition and explain higher risk for cardiovascular events in this group.

Conclusion

Therapist-guided internet-based cognitive behavioral therapy developed for patients with a recent myocardial infarction was not superior to treatment as usual at the 1-year follow-up in the U-CARE Heart trial; however, the CAQ scale did detect a possible trend (*P*<.02, which was not significant when corrected for multiple comparisons) toward improvement in the treatment group. This may indicate a differential effect of the internet-based cognitive behavioral therapy on cardiac anxiety that requires further research. The internet-based cognitive behavioral therapy was not associated with an increased time-to-event risk for cardiovascular events. The low adherence rates should be considered when interpreting these results.

XSL•FO
RenderX

Acknowledgments

This study is part of U-CARE, which is a strategic research environment funded by the Swedish Research Council (2009-1093). The Swedish Heart and Lung Association (E 148/11); the Uppsala-Örebro Regional Research Council (22 86 61); Swedish Research Council for Health, Working Life, and Welfare (2014-4947); and the Vårdal foundation (2014-0114) are also responsible for funding U-CARE Heart. The authors would like to thank all patients who participated in this study for their time and contribution; participating hospitals; nurses and doctors involved who helped them in the recruitment process; Kata Bordas and Malin Fridlund, who helped out with the study recruitment and informing patients; Amanda Romppala, for the administrative help and contribution; and Ronnie Pingel, Maria Bertilsson, and Lars Lindhagen from Uppsala University for statistical support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of the U-CARE Portal showing the U-CARE Heart intervention.
[PNG File , 723 KB - jmir_v23i5e25465_app1.png]

Multimedia Appendix 2
Outcome mean change and treatment effects at baseline, posttreatment, and 1-year follow-up.
[DOCX File , 16 KB - jmir_v23i5e25465_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2136 KB - jmir_v23i5e25465_app3.pdf]

References

1. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
2. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
3. Geiser F, Urbach AS, Harbrecht U, Conrad R, Pötzsch B, Amann N, et al. Anxiety and depression in patients three months after myocardial infarction: association with markers of coagulation and the relevance of age. J Psychosom Res 2017 Aug;99:162-168. [doi: 10.1016/j.jpsychores.2017.06.015] [Medline: 28712423]
4. Norlund F, Lissåker C, Wallert J, Held C, Olsson EM. Factors associated with emotional distress in patients with myocardial infarction: results from the SWEDEHEART registry. Eur J Prev Cardiol 2018 Jun;25(9):910-920 [FREE Full text] [doi: 10.1177/2047487318770510] [Medline: 29692223]
5. Wu Q, Kling JM. Depression and the risk of myocardial infarction and coronary death: a meta-analysis of prospective cohort studies. Medicine (Baltimore) 2016 Feb;95(6):e2815 [FREE Full text] [doi: 10.1097/MD.0000000000002815] [Medline: 26871852]
6. Lesperance F, Frasure-Smith N, Talajic M. Major depression before and after myocardial infarction: its nature and consequences. Psychosom Med 1996;58(2):99-110. [doi: 10.1097/00006842-199603000-00001] [Medline: 8849624]
7. Sundbøll J, Schmidt M, Adelborg K, Pedersen L, Bøtker HE, Videbech P, et al. Impact of pre-admission depression on mortality following myocardial infarction. Br J Psychiatry 2017 May 02;210(5):356-361. [doi: 10.1192/bjp.bp.116.194605] [Medline: 28254961]
8. Smolderen KG, Strait KM, Dreyer RP, D'Onofrio G, Zhou S, Lichtman JH, et al. Depressive symptoms in younger women and men with acute myocardial infarction: insights from the VIRGO study. J Am Heart Assoc 2015 Apr 02;4(4):e001424 [FREE Full text] [doi: 10.1161/JAHA.114.001424] [Medline: 25836055]
9. Lane D, Carroll D, Ring C, Beevers DG, Lip GY. Mortality and quality of life 12 months after myocardial infarction: effects of depression and anxiety. Psychosom Med 2001;63(2):221-230. [doi: 10.1097/00006842-200103000-00005] [Medline: 11292269]
10. Schweikert B, Hunger M, Meisinger C, König HH, Gapp O, Holle R. Quality of life several years after myocardial infarction: comparing the MONICA/KORA registry to the general population. Eur Heart J 2009 Feb 27;30(4):436-443. [doi: 10.1093/eurheartj/ehn509] [Medline: 19019995]

XSL•FO
RenderX

11. Richards SH, Anderson L, Jenkinson CE, Whalley B, Rees K, Davies P, et al. Psychological interventions for coronary heart disease. Cochrane Database Syst Rev 2017 Dec 28;4:CD002902. [doi: 10.1002/14651858.CD002902.pub4] [Medline: 28452408]
12. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
13. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
14. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
15. Lundgren JG, Dahlström, Andersson G, Jaarsma T, Kärner KA, Johansson P. The effect of guided web-based cognitive behavioral therapy on patients with depressive symptoms and heart failure: a pilot randomized controlled trial. J Med Internet Res 2016 Aug 03;18(8):e194 [FREE Full text] [doi: 10.2196/jmir.5556] [Medline: 27489077]
16. Johansson P, Westas M, Andersson G, Alehagen U, Broström A, Jaarsma T, et al. An internet-based cognitive behavioral therapy program adapted to patients with cardiovascular disease and depression: randomized controlled trial. JMIR Ment Health 2019 Oct 03;6(10):e14648 [FREE Full text] [doi: 10.2196/14648] [Medline: 31584000]
17. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013 Mar 26;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
18. Norlund F, Wallin E, Olsson EMG, Wallert J, Burell G, von Essen L, et al. Internet-based cognitive behavioral therapy for symptoms of depression and anxiety among patients with a recent myocardial infarction: the U-CARE heart randomized controlled trial. J Med Internet Res 2018 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
19. Schneider L, Hadjistavropoulos H, Dear B, Titov N. Efficacy of internet-delivered cognitive behavioural therapy following an acute coronary event: a randomized controlled trial. Internet Interv 2020 Sep;21:100324 [FREE Full text] [doi: 10.1016/j.invent.2020.100324] [Medline: 32455120]
20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
21. Norlund F, Olsson EMG, Burell G, Wallin E, Held C. Treatment of depression and anxiety with internet-based cognitive behavior therapy in patients with a recent myocardial infarction (U-CARE Heart): study protocol for a randomized controlled trial. Trials 2015 Apr 11;16:154 [FREE Full text] [doi: 10.1186/s13063-015-0689-y] [Medline: 25873137]
22. Tesio V, Molinaro S, Castelli L, Andriani M, Fanelli A, Marra S, et al. Screening of depression in cardiac patients: is the HADS an effective tool? Int J Cardiol 2014 Mar 15;172(2):e344-e345. [doi: 10.1016/j.ijcard.2013.12.308] [Medline: 24461980]
23. Christensen AV, Dixon JK, Juel K, Ekholm O, Rasmussen TB, Borregaard B, et al. Psychometric properties of the Danish Hospital Anxiety and Depression Scale in patients with cardiac disease: results from the DenHeart survey. Health Qual Life Outcomes 2020 Jan 07;18(1):9 [FREE Full text] [doi: 10.1186/s12955-019-1264-0] [Medline: 31910859]
24. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 2016 Feb 25;16:45 [FREE Full text] [doi: 10.1186/s12888-016-0735-7] [Medline: 26915661]
25. Eifert GH, Thompson RN, Zvolensky MJ, Edwards K, Frazer NL, Haddad JW, et al. The cardiac anxiety questionnaire: development and preliminary validity. Behav Res Ther 2000 Oct;38(10):1039-1053. [doi: 10.1016/s0005-7967(99)00132-1] [Medline: 11004742]
26. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). J Affect Disord 2001 May;64(2-3):203-216. [doi: 10.1016/s0165-0327(00)00242-1] [Medline: 11313087]
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. Registers. Socialstyrelsen. URL: https://www.socialstyrelsen.se/en/statistics-and-data/registers/ [accessed 2021-05-14]
29. ICD-10: International Statistical Classification of Diseases and Related Health Problems. Geneva: World Health Organization; 2004.
30. Jernberg T, Attebring MF, Hambraeus K, Ivert T, James S, Jeppsson A, et al. The Swedish Web-system for enhancement and development of evidence-based care in heart disease evaluated according to recommended therapies (SWEDEHEART). Heart 2010 Oct;96(20):1617-1621. [doi: 10.1136/hrt.2010.198804] [Medline: 20801780]
31. Buuren SV, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]

XSL•FO
RenderX

32. Wallin E, Norlund F, Olsson EMG, Burell G, Held C, Carlsson T. Treatment activity, user satisfaction, and experienced usability of internet-based cognitive behavioral therapy for adults with depression and anxiety after a myocardial infarction: mixed-methods study. J Med Internet Res 2018 Mar 16;20(3):e87 [FREE Full text] [doi: 10.2196/jmir.9690] [Medline: 29549067]
33. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, INTERHEART Investigators. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. Eur Heart J 2008 Apr;29(7):932-940. [doi: 10.1093/eurheartj/ehn018] [Medline: 18334475]
34. Vroman KG, Arthanat S, Lysack C. "Who over 65 is online?" older adults' dispositions toward information communication technology. Comput Human Behav 2015 Feb;43:156-166. [doi: 10.1016/j.chb.2014.10.018]
35. Shah A, Morthland M, Scogin F, Presnell A, DiNapoli E, DeCoster J, et al. Audio and computer cognitive behavioral therapy for depressive symptoms in older adults: a pilot randomized controlled trial. Behav Ther 2018 Nov;49(6):904-916. [doi: 10.1016/j.beth.2018.06.002] [Medline: 30316489]
36. Hauffman A, Alfonsson S, Bill-Axelson A, Bergkvist L, Forslund M, Mattsson S, et al. Cocreated internet-based stepped care for individuals with cancer and concurrent symptoms of anxiety and depression: Results from the U-CARE AdultCan randomized controlled trial. Psychooncology 2020 Dec;29(12):2012-2018 [FREE Full text] [doi: 10.1002/pon.5489] [Medline: 32691455]
37. Romijn G, Batelaan N, Kok R, Koning J, van Balkom A, Titov N, et al. Internet-delivered cognitive behavioral therapy for anxiety disorders in open community versus clinical service recruitment: meta-analysis. J Med Internet Res 2019 Apr 17;21(4):e11706 [FREE Full text] [doi: 10.2196/11706] [Medline: 30994462]
38. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting adherence to internet-Delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the U-CARE heart randomized controlled trial. J Med Internet Res 2018 Dec 10;20(10):e10754 [FREE Full text] [doi: 10.2196/10754] [Medline: 30305255]
39. Gulliksson M, Burell G, Vessby B, Lundin L, Toss H, Svärdsudd K. Randomized controlled trial of cognitive behavioral therapy vs standard treatment to prevent recurrent cardiovascular events in patients with coronary heart disease: Secondary Prevention in Uppsala Primary Health Care project (SUPRIM). Arch Intern Med 2011 Jan 24;171(2):134-140. [doi: 10.1001/archinternmed.2010.510] [Medline: 21263103]
40. Rafanelli C, Gostoli S, Buzzichelli S, Guidi J, Sirri L, Gallo P, et al. Sequential combination of cognitive-behavioral treatment and well-being therapy in depressed patients with acute coronary syndromes: a randomized controlled trial (TREATED-ACS study). Psychother Psychosom 2020 Aug 13;89(6):345-356 [FREE Full text] [doi: 10.1159/000510006] [Medline: 32791501]
41. Frasure-Smith N, Lespérance F, Prince RH, Verrier P, Garber RA, Juneau M, et al. Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction. Lancet 1997 Aug 16;350(9076):473-479. [doi: 10.1016/S0140-6736(97)02142-9] [Medline: 9274583]
42. Hanssen TA, Nordrehaug JE, Eide GE, Bjelland I, Rokne B. Anxiety and depression after acute myocardial infarction: an 18-month follow-up study with repeated measures and comparison with a reference population. Eur J Cardiovasc Prev Rehabil 2009 Dec;16(6):651-659. [doi: 10.1097/HJR.0b013e32832e4206] [Medline: 19707149]
43. Christina C. An observational study of the occurrence of anxiety, depression and self-reported quality of life 2 years after myocardial infarction. J Cardiol Cardiovasc Med 2018;3(3):052-063. [doi: 10.29328/journal.jccm.1001027]
44. van Beek MHCT, Mingels M, Voshaar RCO, van Balkom AJLM, Lappenschaar M, Pop G, et al. One-year follow up of cardiac anxiety after a myocardial infarction: a latent class analysis. J Psychosom Res 2012 Nov;73(5):362-368. [doi: 10.1016/j.jpsychores.2012.09.004] [Medline: 23062810]

Abbreviations

BADS-SF: Behavioral Activation for Depression Scale–Short Form
CAQ: Cardiac Anxiety Questionnaire
HADS: Hospital Anxiety and Depression Scale
HR: hazard ratio
ICD: International Statistical Classification of Diseases
MADRS-S: Montgomery-Åsberg Depression Scale–Self-rated
SMS: short message service

XSL•FO
RenderX

Edited by G Eysenbach; submitted 03.11.20; peer-reviewed by M Schure, E Gustafson; comments to author 20.11.20; revised version received 24.11.20; accepted 26.04.21; published 24.05.21.

Please cite as:
Humphries SM, Wallert J, Norlund F, Wallin E, Burell G, von Essen L, Held C, Olsson EMG
Internet-Based Cognitive Behavioral Therapy for Patients Reporting Symptoms of Anxiety and Depression After Myocardial Infarction: U-CARE Heart Randomized Controlled Trial Twelve-Month Follow-up
J Med Internet Res 2021;23(5):e25465
URL: https://www.jmir.org/2021/5/e25465
doi:10.2196/25465
PMID:34028358

©Sophia Monica Humphries, John Wallert, Fredrika Norlund, Emma Wallin, Gunilla Burell, Louise von Essen, Claes Held, Erik Martin Gustaf Olsson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.05.2021. 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

Intelligent Personalized Exercise Prescription Based on an eHealth Promotion System to Improve Health Outcomes of Middle-Aged and Older Adult Community Dwellers: Pretest–Posttest Study

Ting Sun[1], MPhil; Yang Xu[2], PhD; Hui Xie[1], PhD; Zuchang Ma[2], PhD; Yu Wang[2], MPhil

[1]School of Nursing, Bengbu Medical College, Bengbu, China

[2]Anhui Province Key Laboratory of Medical Physics and Technology, Institute of Intelligent Machines, Hefei Institutes of Physical Sciences, Chinese Academy of Sciences, Hefei, China

Corresponding Author:
Zuchang Ma, PhD
Anhui Province Key Laboratory of Medical Physics and Technology
Institute of Intelligent Machines
Hefei Institutes of Physical Sciences, Chinese Academy of Sciences
350# Shushan Lake Rode
Hefei, 230031
China
Phone: 86 0551 65591104
Email: ZCMa121@126.com

Abstract

Background: A scientific, personalized, and quantitative exercise prescription that has the potential to be an important therapeutic agent for all ages in the prevention of chronic disease is highly recommended. However, it is often poorly implemented, as clinicians lack the necessary knowledge and skills while participants have low adherence due to design defects (eg, prescriptions fail to take individual willingness, the appeal of exercise, and complex physical conditions into account). Intelligent personalized prescription is thus worth exploring.

Objective: The aim of this study was to investigate whether a year-long cloud platform–based and intelligent personalized exercise prescription intervention could improve Chinese middle-aged and older adult community dwellers' health outcomes.

Methods: A total of 177 participants (aged 52-85 years; mean 67.93, SD 7.05) were recruited from 2 Chinese community health service centers in Anhui Province, China. The exercise intervention was delivered over 12 months with a single-group pretest–posttest design. After being assessed in terms of physical activity, health-related lifestyle, history of chronic diseases and drug use, family history of disease and cardiovascular function, body composition, bone mineral density, and physical fitness through an eHealth promotion system, participants with relative contraindications for exercise were personally prescribed the health care exercise mode by an intelligent system, while those without relative contraindication and who had a regular exercise habit were prescribed the scientific fitness mode. Paired *t* tests were used for the analysis.

Results: A total of 97 participants were classified into the health care mode, and the remaining 80 participants were assigned to the scientific fitness mode. Significant changes in heart rate (mean difference [MD] 2.97; 95% CI 1.1-4.84; *P*=.002), subendocardial viability ratio (MD –0.13; CI: –1.19 to –0.63; *P*<.001), weight (MD 0.99; CI 0.29-1.69; *P*=.006), BMI (MD 0.38; CI 0.11-0.64; *P*=.006), body fat rate (MD 0.88; CI 0.24-1.51; *P*=.007), fat mass (MD 0.92; CI 0.33-1.53; *P*=.003), and brachial-ankle pulse wave velocity (MD: –0.72; CI –1.17 to –0.27; *P*=.002) were observed among participants with the health care mode exercise prescriptions at the 12-month postintervention versus the baseline assessment, while no changes in systolic blood pressure, diastolic blood pressure, muscle mass, bone mineral density, *t* value, *z* value, balance, or ability were discerned. The results showed a functional decline in the physical fitness of both groups, including in handgrip strength (healthcare mode: MD 4.41; scientific fitness mode: MD 3.11), vital capacity (healthcare mode: MD 261.99; scientific fitness mode: MD 250.78), and agility (healthcare mode MD=–0.35; scientific fitness mode: MD=–0.39) with all *P* values <.001, except handgrip strength in the scientific fitness mode (*P*=.002). There were no significant differences in other parameters among participants with scientific fitness mode exercise prescriptions.

XSL•FO
RenderX

Conclusions: The observations suggest that our exercise prescription intervention program might promote certain health outcomes in Chinese middle-aged and older adult community dwellers, yet we are unable to recommend such a program given the existing limitations. Future randomized controlled trials with diverse samples are warranted to confirm our findings.

(J Med Internet Res 2021;23(5):e28221) doi:10.2196/28221

KEYWORDS

exercise prescription; cardiovascular function; body composition; bone mineral density; physical fitness

Introduction

Physical activity is defined as any bodily movement produced by skeletal muscle [1]. Exercise is one form of delivery for physical activity and has been defined as a structured activity to maintain essential physiological systems, such as the skeletal muscular and metabolic systems [2], and to improve physical function and quality of life [1,3-5]. Resistance training is considered a very effective method for the development of skeletal muscle [6], and aerobic exercise training performed at an appropriate level of intensity has beneficial effects on cardiopulmonary function [7]. As middle-aged and older adults experience functional decline and often suffer from one or more noncommunicable chronic diseases (NCDs), many people have become aware of the health benefits of regular exercise. Consistent evidence confirms the benefits of exercise: prevention of injurious falls, sarcopenia, cognitive decline, and frailty [3,8,9]; a reduction of morbidity (eg, coronary heart disease, type 2 diabetes) [2,3,10,11]; and at least a 20%-30% reduction for more than 25 chronic symptoms (eg, depression or hypertension,) and 10% reduction in premature mortality [10,12]. Furthermore, a lack of exercise not only leads to an increase in NCDs, but also increases medical costs and the economic burden on the state and individuals [4], which is estimated to be at about US $68 billion worldwide annually [10]. The benefits of exercise for everyone are obvious. Even for middle-aged and older adults with NCDs, exercise is one of the best medicines, being cheap, accessible, and providing a complex, whole-body impact with very few side effects [13]. Therefore, exercise training is a cornerstone in the management of NCDs [14].

Although the value of regular exercise is well known, many people still do not include exercise in their daily routines [15]. A global pandemic of physical inactivity has been described [10]. For example, 50% of American adults and 49% of New Zealand adults are nonactive or insufficiently active [2,12]. The situation for older people may be worse as a result of complex physical conditions [16]. Older individuals may struggle to exercise because of pain or other symptoms [17].

Exercise prescription is cost-effective and can increase physical activity by 10% in relatively inactive patients [1]. It is recommended and defined by the World Health Organization (WHO) as follows: an exercise prescription given by physicians to help patients engage in physical exercise, according to the medical examination data (including exercise and physical tests) and based on the patient's specific health, physical strength, and cardiovascular function status. The type of exercise, exercise intensity, exercise time, and exercise frequency are specified in the prescription, and any associated precautions when undertaking the exercise are detailed [18]. The American College of Sports Medicine (ACSM) introduced an exercise prescription guideline in 1990 to guide the design of exercise prescriptions for the general population and people with chronic diseases [19].

Exercise prescription is highly recommended by the current guidelines on the prevention and management of chronic disease, but its implementation remains poor [20]. Many clinicians—especially primary care staff [1]—experience difficulties in prescribing exercise in the presence of different concomitant chronic diseases and risk factors within the same patient because of a lack of expert knowledge and skills [21,22]. Although the scientific effectiveness of exercise prescriptions is supported by a substantial amount of experimental data, studies have also demonstrated that adherence to this "scientific" and "best health–benefit" exercise prescription is not high [23,24]. This is likely because the prescription design does not take into account factors such as individual willingness to exercise, the appeal of exercise, and the conditions under which exercise occurs. In short, personalized exercise prescription is not simple and comprises many challenges, such as time constraints, complex comorbidities, perceived lack of patient engagement, and a lack of physician training or education on the particulars of physical activity [1,13]. To provide an effective and safe exercise program for patients with multiple comorbidities, an advanced intelligent exercise prescription that can calculate the complex physical condition of a patient and combine this with the patient's exercise preferences to give the best prescription, is required [25]

The purpose of this year-long program was to verify the effectiveness of intelligent personalized exercise prescriptions for middle-aged and older adult community dwellers and support the next research stage: exercise prescription using the smartphone with intervention and procedural monitoring through wearable devices. Ultimately, we envision a new protocol that includes 3 characteristics: (1) an individual assessment of needs, motivation, habits, preferences, and barriers; (2) valid behavior change approaches; and (3) proper follow-up, self-monitoring, and social support.

Methods

Design

This study included a single-group pretest–posttest design with blind pre- and postoutcome assessment and the collector of participant information being unaware of the purpose of the experiment.

XSL•FO
RenderX

Study Sample and Recruitment

Between January 2019 and November 2019, middle-aged and older adult community-dwelling individuals were recruited via 2 community health service centers in Bengbu in the Anhui Province of China. They were invited by center staff through telephone or verbal invitation, and those interested were asked to contact staff members for further information. Eligible participants were then scheduled to be pre-examined one by one. When they finished the entire year-long intervention program, the time of the postintervention evaluations was booked, and they were reassessed one by one until the last participant finished in November 2020. There were 3 modes available for the intervention, but because of difficulties in meeting requirements for the "exercise habit formation" mode, we only recruited participants for the "health care" and "scientific fitness mode" groups. The participants were required to be more than 50 years of age, be community dwellers, have no regular exercise habits (total number of physical activities of various intensity per week ≥3 times; total physical activity time per week ≥90 minutes) for inclusion into the health care mode group, and have regular exercise habits for inclusion into the scientific fitness mode group. Participants were excluded if they had serious cardiovascular or cerebrovascular diseases; lung or kidney diseases or related complications, according to their self-reported hospitalization experience; severe diabetes or related complications, such as fundus lesion, peripheral neuropathy, diabetic foot, or renal dysfunction; a fasting blood glucose ≥13.3 mmol/L and urine ketone positive, or postprandial blood glucose ≥19.4 mmol/, with resting blood pressure ≥180/110 mmHg; or severe cognitive impairment and diagnosed mental disorders, such as moderate to severe depression, schizophrenia, and mania.

Ethical Considerations

In accordance with ethical guidelines, the participants were fully informed about the positive effects of exercise prescription and the details of the procedures. All participants provided written informed consent to participate in this study and agreed to their data being used. The study protocol was approved by the Ethics Committee of Bengbu Medical College (Anhui, China; no. 2018045).

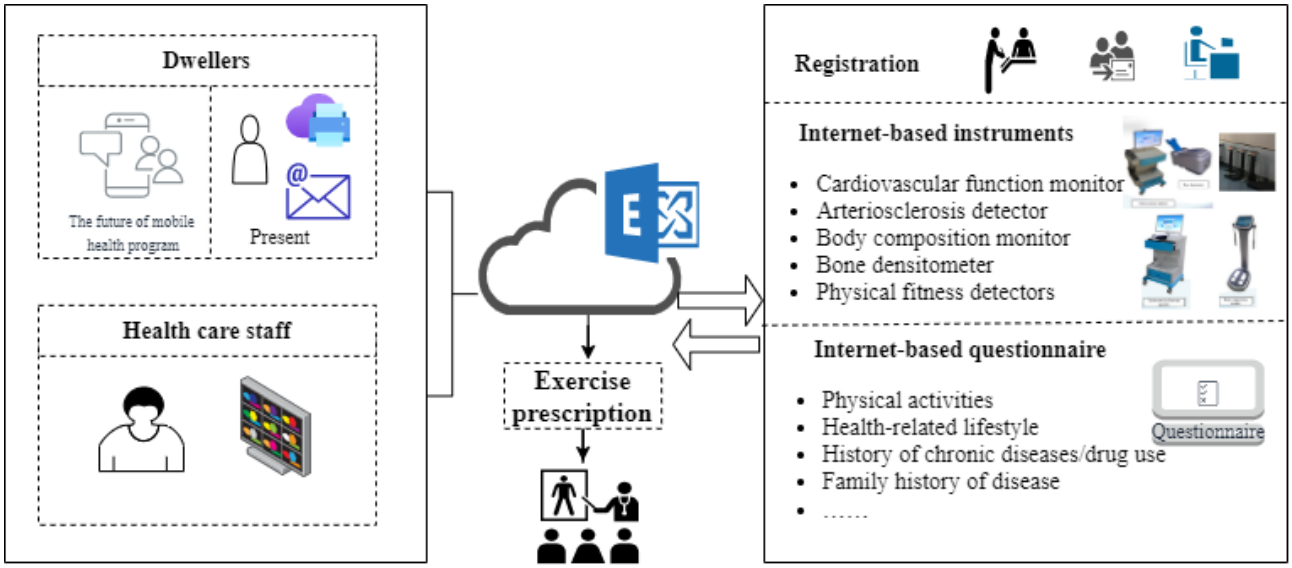
Sample Size

We calculated the sample size of the mean difference (MD) of the before and after comparison of each index, as reported in similar literature [26], using MedSci sample size tools with α=.05 and 1-β=.8 resulting in a sample range of 11-67. A sample size of 70 was decided upon for both groups.

Intervention

This intelligent personalized exercise prescription program (IPEPP) is a preliminary experiment for a mobile health intervention, based on an eHealth promotion system (Figure 1). At present, it includes 4 elements: (1) registration system (typically participants are invited to participate in the program by a family-contracted physician); (2) cloud platform (all data are stored and calculated in the cloud platform, through which community health care staff and researchers can monitor an individual's data online); (3) internet-based instruments (including cardiovascular function monitor, arteriosclerosis detector, body composition monitor, bone densitometer, and physical fitness detectors used to assess cardiovascular function, body composition, bone mineral density, and physical fitness of participants before and after intervention); (4) an internet-based questionnaire to collect data about physical activity, health-related lifestyle, history of chronic diseases or drug use, and family history of disease. In the future, a mobile health app currently under development will be combined with wearable devices for monitoring the process of exercise to increase the adherence of intervention based on theories of health behavior change.

Figure 1. Components of the eHealth promotion system.



The IPEPP comprised 4 processes: (1) assessment (using internet-based instruments and an internet-based questionnaire to collect integrated information and analyze the situation of the individuals to support decision-making about individualized exercise prescription), (2) exercise prescription generation (based on the assessment of the integrated information and the

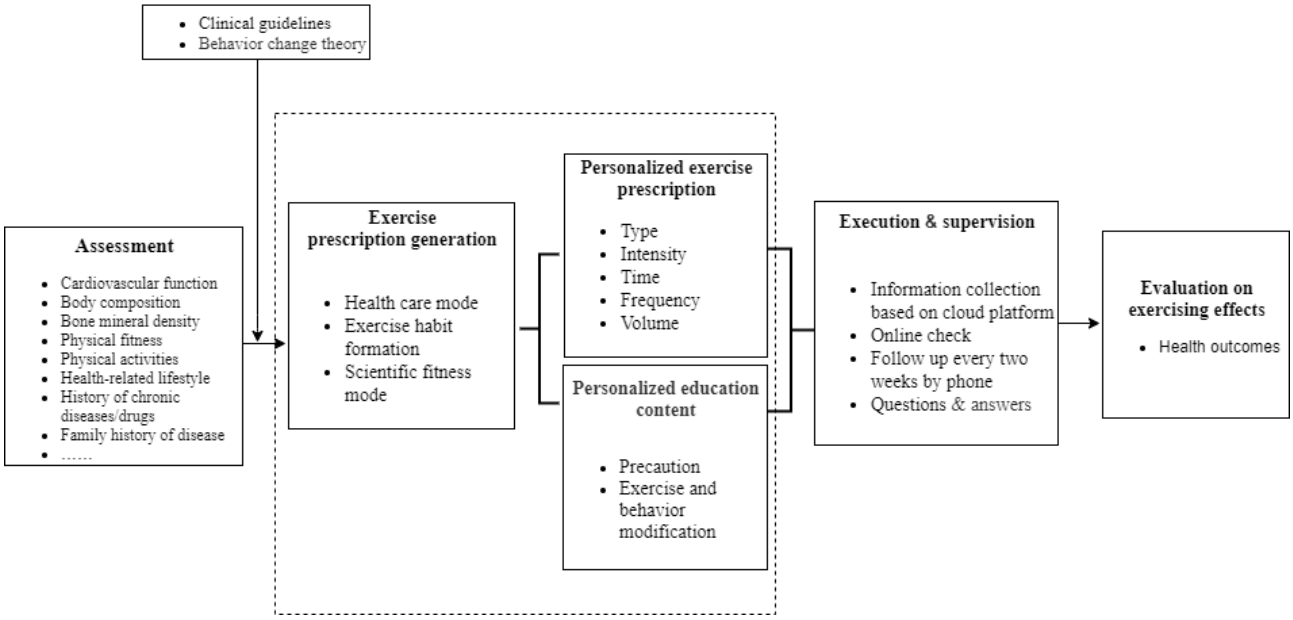
XSL•FO
RenderX

individual's exercise habits, 3 types of personalized exercise prescription are generated: a health care mode, an exercise habit-formation mode, and a scientific fitness mode), (3) execution and supervision (health care staff print exercise prescriptions or send them to participants by email, explain the instructions, precautions, and content of exercise and behavior modification, and follow up with participants every 2 weeks by telephone), and (4) evaluation of exercising effects (cardiovascular function, body composition, bone mineral density, and physical fitness).

Assessment

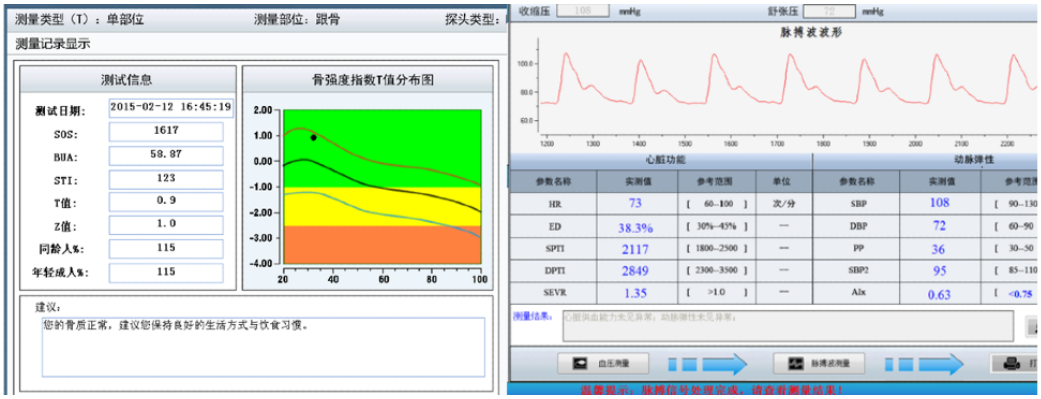
According to the ACSM's guidelines for exercise testing and prescription, pre-exercise screening is indispensable [19]. The generation of our exercise prescription was based on the assessment of baseline data. It consisted of information collected by an internet-based questionnaire system, including the individual's current level of physical activity, history of chronic diseases (eg, cardiovascular and cerebrovascular diseases, diabetes and hypertension) and drug use (eg, types, dosage, and times), symptoms of the disease (eg, exercise dysfunction, hypoglycemia, and hypotension symptoms), health-related lifestyle (eg, eating, smoking and drinking habits), and family history of disease (Figure 2; other details are published elsewhere [27]); and information collected by internet-based instruments, which included physiological parameters (cardiovascular function, body composition, bone mineral density, and physical fitness) as obtained from a body composition monitor (BX-BCA-100, Institute of Intelligent Machines), a cardiovascular function monitor (BX-CFTI-100, Institute of Intelligent Machines), an arteriosclerosis detector (BX-AS-100, Institute of Intelligent Machines), a bone densitometer (BX-BDI-500A, Institute of Intelligent Machines), and physical fitness detectors, including a handgrip strength meter (TSN100-WL, Physical Fitness Sports Technology Company), a reaction time meter (TSN100-FY, Physical Fitness Sports Technology Company), a 1-leg stand meter with closed eyes (TSN100-ZL, Physical Fitness Sports Technology Company), a spirometer (TSN100-FH, Physical Fitness Sports Technology Company), and a flexion measurement instrument of sitting position (TSN100-TQ, Physical Fitness Sports Technology Company). Figure 3 displays examples of screenshots showing examination results of the bone densitometer and cardiovascular function monitor. The details of the assessment procedure are shown in Multimedia Appendix 1, and other related information has been published elsewhere[27].

Figure 2. The assessment, exercise prescription generation, execution and supervision, and evaluation of exercising effects of the intelligent personalized exercise prescription program intervention.



XSL•FO
RenderX

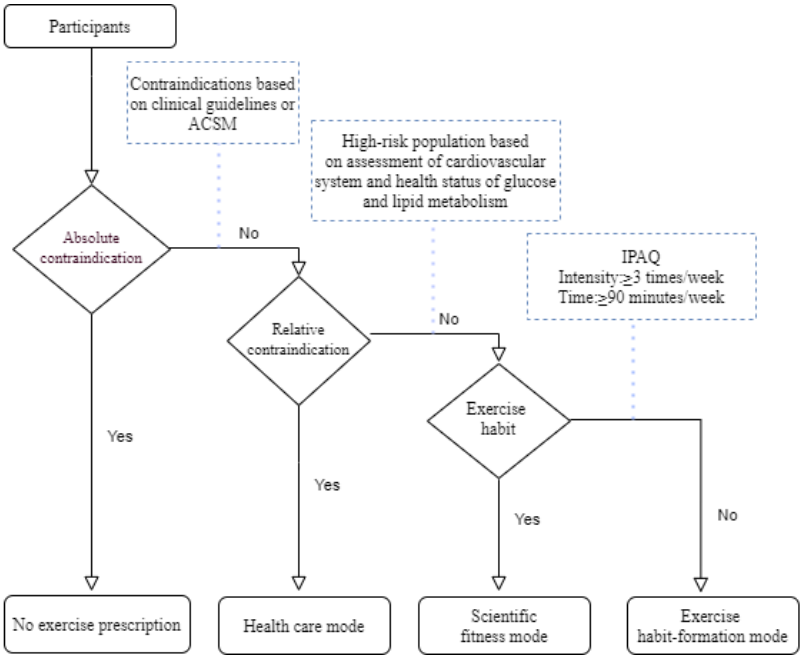
Figure 3. Examples of screenshots of the examination reports of bone mineral density and cardiovascular function.



Exercise Prescription Generation

All exercise prescription recommendations were based on clinical guidelines and the ACSM. Participants for whom exercise was absolutely contraindicated were excluded, while those with relative contraindications were given the health care mode as their prescription. Relative contraindications included the following: suspected symptoms or target organ damage, as indicated by subendocardial viability ratio (SEVR) ≤0.9; ankle/brachial systolic pressure index (ABI) ≥1.4 or 0.6 ≥ ABI >0, brachial-ankle pulse wave velocity (baPWV) ≥21.1 m/s; central arterial pressure ≥145 mmHg; and abnormal blood pressure, as indicated by systolic blood pressure (SBP) ≥160 mmHg or diastolic blood pressure (DBP) ≥100 mmHg, and SBP <90 mmHg or DBP <60 mmHg with hypotensive symptoms. After being confirming of having no relative contraindications, participants with or without adequate and regular physical activities were given the scientific fitness mode or exercise habit-formation mode exercise prescriptions (Figure 4).

Figure 4. Algorithms for 3 modes of exercise prescription. ACSM: The American College of Sports Medicine. IPAQ: International Physical Activity Questionnaire.



The system automatically generated 3 kinds of exercise prescriptions with pictures of the recommended exercise type (Figure 5) and notes of exercise precautions corresponding with the screened risks and incorrect exercise-related health behaviors. The health care mode recommended low-intensity aerobic gymnastics and other simple types of low-intensity exercise, including basic self-massage exercises (randomly drawn from the system). Compared with the health care mode, the exercise habit-formation mode recommended an unstructured prescription to cultivate individuals' exercise consciousness. It encouraged individuals to choose their exercise type and intensity according to their subjective feelings, and exercises were recommended with low requirements for skills or physical fitness, such as walking, leisure cycling, and slow dances. It is generally recommended that a single effective exercise time should not be less than 10 minutes and that the frequency of exercise should not be less than twice a week. The scientific fitness mode recommended a structured prescription following the frequency, intensity, type, time, volume, and progression principle, which includes aerobic exercise and resistance training (Table 1). Examples of exercise prescriptions and algorithms for various exercise categories are presented in Multimedia Appendices 2 and 3.

XSL•FO
RenderX

Figure 5. Examples of exercise instructions.

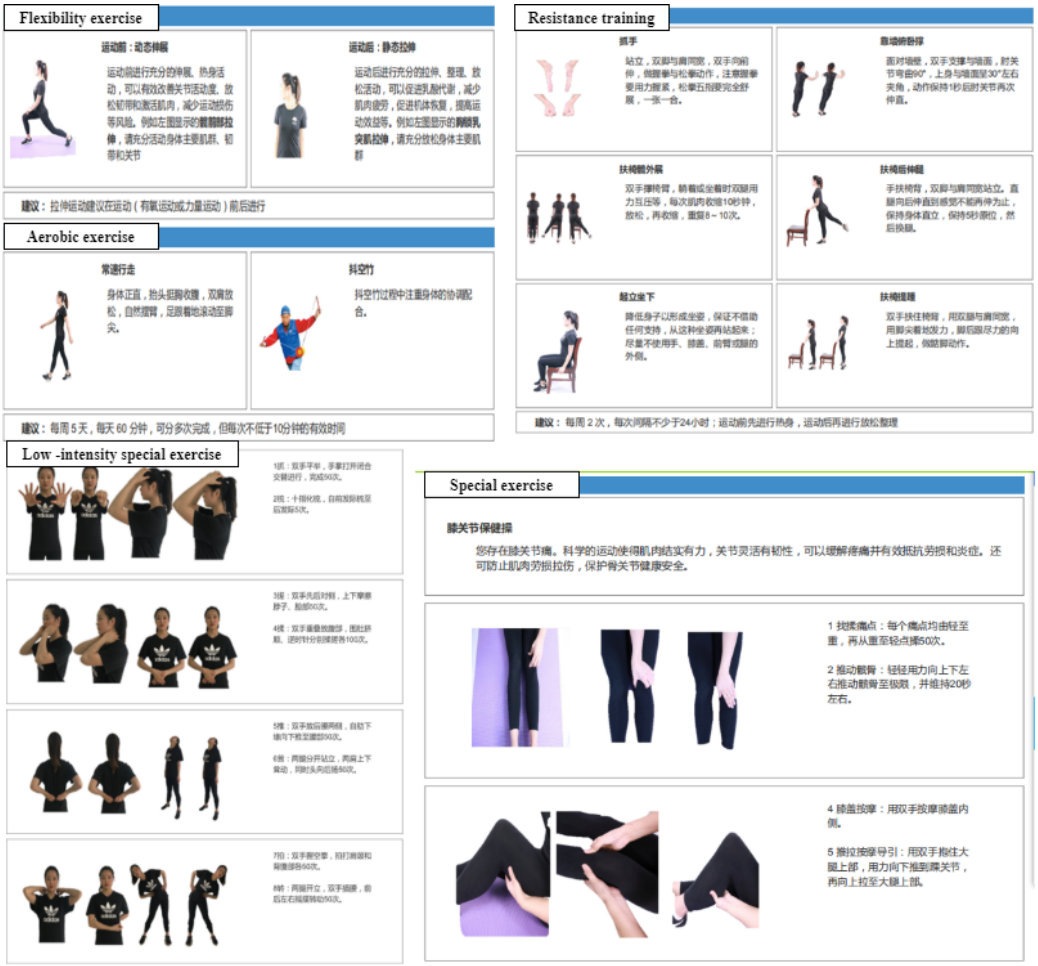


Table 1. Matrix of links between the categories of personalized exercise prescriptions and exercise types.

Prescription type	Aerobic exercise	Resistance training	Flexibility exercise	Special exercise[a]	Low-intensity aerobic gymnastics
Health care mode	√[b]				√
Exercise habit-formation mode	√		√	√	
Scientific fitness mode	√	√	√	√	

[a]Special exercise is recommended by prescription when a related health problem is detected in the exercise habit-formation mode and scientific fitness mode.

[b]√: The type of exercise recommended by prescription.

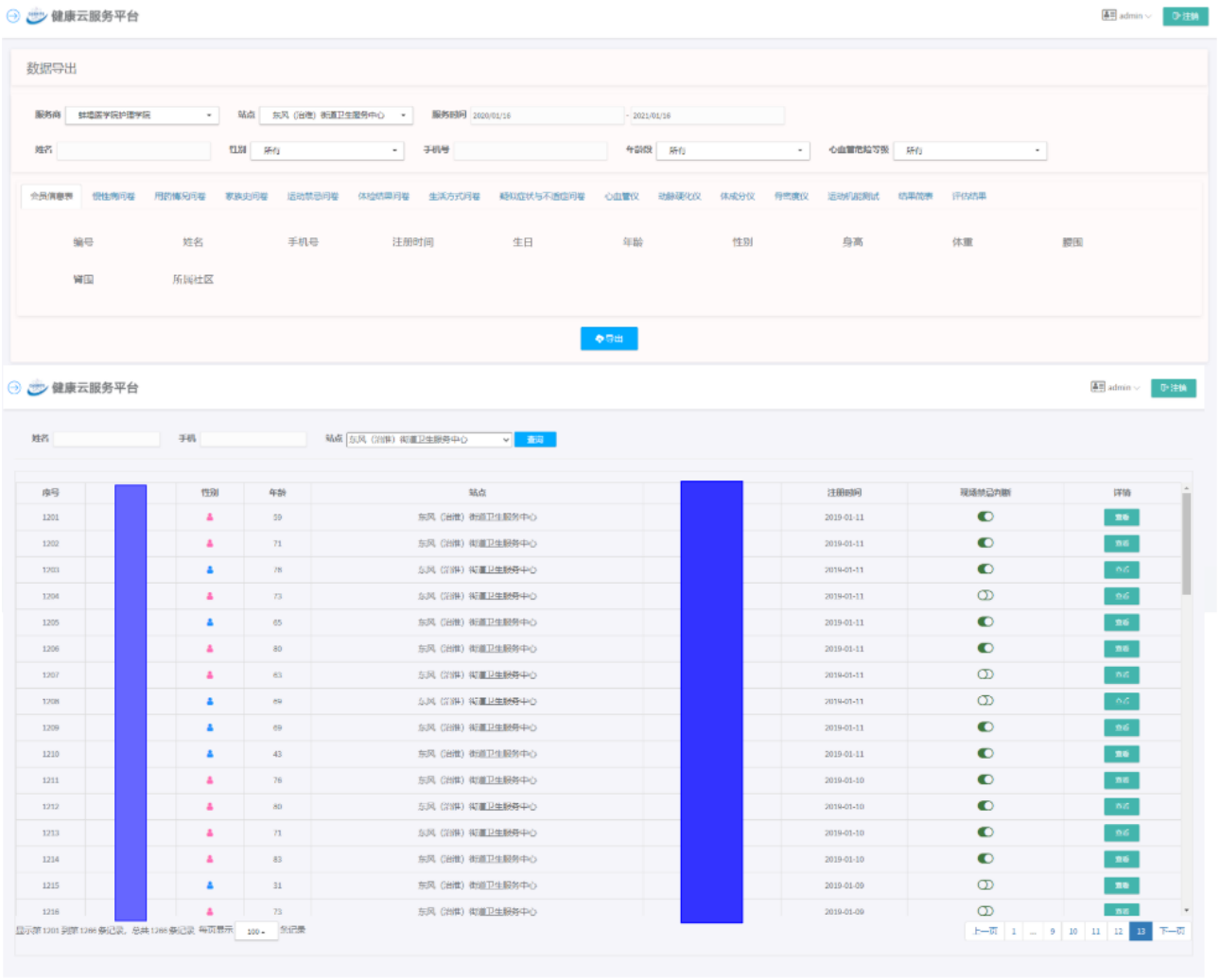
The exercise habit-formation mode and scientific fitness mode include recommendations for flexibility exercises before and after either aerobic exercise or resistance training (Figure 5). Special exercises were recommended for any currently detected health problems (eg, knee pain, neck discomfort), and the system recommended a suitable set of health care exercises from the special sports library as guidance. Only when relevant problems were detected in the participants during the assessment process would these special exercises be recommended (Table 1).

Execution, Supervision, and Evaluation of Exercising Effects

Participants were given their exercise prescription in the community health center after a detailed explanation, which culminated in them taking away printed color papers or receiving emails. Researchers visited the cloud platform online to find each individual's information (Figure 6), and they followed up participants every 2 weeks by telephone to remind them to exercise according to their prescription (keeping the same prescription for a year) and to help them solve any problems encountered.

XSL•FO
RenderX

Figure 6. Website with cloud platform screenshot.



After 12 months, the participants were reevaluated to check if there were any changes in the following health outcomes: cardiovascular function, body composition, bone mineral density, and physical fitness. Participants were asked to report whether they changed their medication use or related lifestyle (eg, diet change).

Statistical Analyses

Data were analyzed using SPSS 23.0 software (IBM Corp). Continuous variables are expressed as mean and SD. Paired *t* tests for independent samples and Pearson chi-squared test were used to assess the significance of differences in baseline characteristics between the health care mode and scientific fitness mode. Paired *t* tests were also used to analyze the difference in health outcomes before and after intervention. A *P* value <0.05 was considered statistically significant.

Results

Participants

A total of 232 participants agreed to participate in the study. Of this number, 27 participants were excluded due to not meeting inclusion criteria, and 28 participants did not complete the entire program including those who changed medication. These 28 participants were excluded due to inadequate adherence, an inability to finish the postassessment, changes in health status unrelated to the study and thus cessation of the intervention, withdrawal, or loss of contact. Overall, 177 participants (105 women and 72 men; mean age 67.93 years, SD 7.05 years) were included in the final analysis, with 97 in the health care mode group and 80 in the scientific fitness mode group. All participants reported they did not change their related lifestyle.

Baseline Data

Baseline characteristics are presented in Table 2. Epidemiological data of the 2 groups showed no difference except for gender, but participants with an exercise prescription in the health care mode had a poorer baseline condition for heart rate, SEVR, baPWV, body fat rate, fat mass, and vital capacity compared with those in the scientific fitness mode group. Other baseline characteristics did not differ between participants allocated to the 2 exercise prescriptions.

XSL•FO
RenderX

Table 2. Baseline data for health care mode versus the scientific fitness mode.

Characteristic	Health care mode (n=97)	Scientific fitness mode (n=80)	*P* value
Epidemiological data			
Male gender, n (%)	30 (30.9)	42 (52.5)	.003
Age (years), mean (SD)	67.76 (6.67)	68.09 (7.42)	.76
Weight (kg), mean (SD)	68.18 (11.55)	68.06 (11.8)	.95
Cigarette consumption, (cigarettes/lifetime), mean (SD)	32022.16 (9934.85)	45579.37 (11378.80)	.37
Alcohol consumption (g/day), mean (SD)	5.97 (15.97)	5.63 (17.26)	.89
Sedentary time (min/day), mean (SD)	74.62 (86.04)	83.42 (85.35)	.51
Diabetes, n (%)	27 (27.8)	18 (22.5)	.27
Hypertension, n (%)	56 (57.7)	36 (45)	.06
Cardiovascular diseases, n (%)	18 (18.6)	15 (18.8)	.56
Cardiovascular function, mean (SD)			
SBP[a] (mmHg)	130.22 (16.06)	125.96 (14.66)	.72
DBP[b] (mmHg)	73.20 (9.10)	74.64 (8.73)	.29
Heart rate (bp/min)	71.66 (10.21)	67.77 (8.23)	.007
SEVR[c]	1.07 (0.19)	1.19 (0.20)	<.001
baPWV[d] (m/s)	17.31 (3.19)	15.70 (2.70)	<.001
Body composition, mean (SD)			
Weight (kg)	68.18 (11.55)	68.06 (11.80)	.95
BMI (kg/m^2)	25.82 (3.99)	25.17 (3.64)	.26
Body-fat rate (%)	30.50 (9.09)	27.03 (8.44)	.01
Fat free mass (kg)	47.18 (8.21)	49.56 (9.04)	.07
Muscle mass (kg)	44.57 (7.90)	46.90 (8.70)	.06
Fat mass (kg)	21.11 (8.67)	18.68 (7.28)	.046
Bone mineral density, mean (SD)			
STI[e]	84.41 (17.66)	90.63 (23.59)	.07
t value[f]	–1.13 (0.93)	–0.82 (1.24)	.08
z value[g]	0.46 (1.41)	0.94 (1.88)	.08
Physical fitness, mean (SD)			
Handgrip strength (kg)	25.01 (7.38)	26.65 (8.46)	.17
Vital capacity (ml)	1776.29 (579.87)	1947.74 (469.06)	.03
Agility (s)	0.89 (0.56)	0.76 (0.29)	.07
Balance ability (s)	4.78 (6.13)	5.29 (7.56)	.63
Flexibility (cm)	7.44 (6.73)	8.21 (6.40)	.44

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]SEVR: subendocardial viability ratio.

[d]baPWV: brachial-ankle pulse wave velocity.

[e]STI: stiffness index.

[f]Used to evaluate the absolute risk of fracture.

[g]Primarily used to assess the relative risk of fracture for comparison between peers.

XSL•FO
RenderX

Treatment Effect

Results of the health care mode group at baseline and posttest on health outcomes are shown in Table 3. Some domains of cardiovascular function and body composition, including heart rate, SEVR, weight, BMI, body fat rate, and fat mass, showed significant improvement. Compared with the scientific fitness mode group (Multimedia Appendix 4), no significant changes were observed in these parameters in the health care mode group. Noticeably, baPWV (MD –0.72; CI –1.17 to –0.27; *P*=.002) changed from 17.31 m/s to 18.03 m/s during the intervention periods in the health care mode participants, while no change was discerned in those following the scientific fitness mode.

For bone mineral density, no significant improvements were shown in either the health care (*P*=.682) or scientific fitness mode groups (*P*=.55). The results showed functional decline in the physical fitness of both groups, including handgrip strength (healthcare mode MD 4.41; scientific fitness mode: MD 3.11), vital capacity (healthcare mode: MD 261.99; scientific fitness mode: MD 250.78), and agility (healthcare mode: MD –0.35; scientific fitness mode: MD –0.39) with all *P* values <.001 except that of handgrip strength in the scientific fitness mode (*P*=.002).

Table 3. Changes of health outcomes in healthcare mode (N=97).

Characteristic	Pretest, mean (SD)	Posttest, mean (SD)	Difference, mean (95% CI)	*P* value
Cardiovascular function				
SBP[a] (mmHg)	130.44 (15.99)	132.41 (14.96)	–1.97 (–4.68 to 0.75)	.15
DBP[b] (mmHg)	73.29 (9.10)	74.54 (9.23)	–1.25 (–2.92 to 0.42)	.14
Heart rate, (bp/min)	72.90 (10.45)	69.93 (8.53)	2.97 (1.1 to 4.84)	.002
SEVR[c]	1.07 (0.20)	1.19 (0.28)	–0.13 (–0.19 to –0.06)	<.001
baPWV[d] (m/s)	17.31 (3.19)	18.03 (2.92)	–0.72 (–1.17 to –0.27)	.002
Body composition				
Weight (kg)	68.18 (11.55)	67.18 (10.83)	0.99 (0.29 to 1.69)	.006
BMI (kg/m^2)	25.85 (3.99)	25.45 (3.65)	0.38 (0.11 to 0.64)	.006
Body fat rate (%)	30.39 (9.13)	29.51 (8.88)	0.88 (0.24 to 1.51)	.007
Fat-free mass (kg)	47.32 (8.22)	47.11 (8.14)	0.21 (–0.16 to 0.58)	.26
Muscle mass (kg)	44.70 (7.92)	44.52 (7.84)	0.19 (–0.15 to 0.52)	.28
Fat mass (kg)	21.15 (8.66)	20.15 (8.00)	0.92 (0.33 to 1.53)	.003
Bone mineral density				
STI[e]	84.41 (17.66)	83.41 (19.56)	1.00 (–3.84 to 5.84)	.68
t value[f]	–1.13 (0.93)	–1.19 (1.04)	0.53 (–0.21 to 0.31)	.69
z value[g]	0.46 (1.41)	0.41 (1.53)	0.54 (–0.35 to 0.45)	.79
Physical fitness				
Handgrip strength (kg)	25.01 (7.38)	20.60 (8.14)	4.41 (2.96, 5.86)	<.001
Vital capacity (ml)	1776.29 (579.87)	1514.00 (568.43)	261.99 (143.84 to 380.14)	<.001
Agility (s)	0.89 (0.56)	1.23 (0.80)	–0.35 (–0.53 to –0.17)	<.001
Balance ability (s)	4.78 (6.13)	6.46 (8.64)	–1.68 (–3.54 to 0.17)	.08
Flexibility (cm)	7.44 (6.73)	8.53 (7.15)	–1.08 (–2.68 to 0.51)	.18

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]SEVR: subendocardial viability ratio.

[d]baPWV: brachial-ankle pulse wave velocity.

[e]STI: stiffness index.

[f]Used to evaluate the absolute risk of fracture.

[g]Primarily used to assess the relative risk of fracture for comparison between peers.

XSL•FO
RenderX

Discussion

We found a few improvements in cardiovascular function and body composition of participants with exercise-related contraindications but without good exercise habits. However, for participants already in the habit of exercising, there were no changes in these 2 domains.

Cardiovascular function and body composition improvements were important outcomes in our study. For cardiovascular function, we detected that heart rate and SEVR were significantly improved following a 12-month unstructured exercise prescription intervention among middle-aged and older adult dwellers with related exercise contraindications. This is consistent with previous studies, in which exercise improved cardiovascular and arterial function [5] for healthy populations, and even sedentary individuals aged more than 50 years gained the benefit of improved maximal oxygen consumption after exercise [28]. Few studies have reported on improvement in SEVR among older adults, but Huang et al [29] found an augmentation in SEVR after a 6-week exercise intervention in obese adolescents. We suspect our low-intensity, unstructured exercise prescription with associated precautions and tips was a major contributor to adherence of participants who have doubts about how to exercise because of their disease symptoms. Our recommended low-intensity special aerobic gymnastics with instruction through pictures gave participants guidelines, and the unstructured prescription provided tools and cues for participants to take safety into consideration while choosing the physical activity they prefer. For body composition, those participants prescribed the health care mode changed their weight, BMI, body fat rate, and fat mass significantly, but their muscle mass was unchanged. Unsurprisingly, this is in line with results of other interventions with or without calory restriction, implemented in a variety of populations [25-30]. Most of these were short-term intervention programs (6 to 12 weeks), and evidence of long-term effectiveness of exercise intervention in middle-aged and older adults is insufficient. It can therefore be concluded that even low-intensity, aerobic exercise and unstructured physical activity is beneficial in the long-term for middle-aged and older adults without exercise habits. Essentially, starting to exercise is the most important thing.

Corresponding parameters of cardiovascular function and body composition in the scientific fitness mode group showed no statistically significant changes. We expected that the structured and combined exercise of aerobic and resistance training would improve cardiovascular function and body composition in participants with exercise habits. However, a possible ceiling effect might have prevented further improvements in functioning, even though these participants had completed more scientific exercise [17]. Theresa et al [30] reported that individuals with more “favorable” values at baseline (eg, lower submaximal heart rate) may potentially show a low subsequent training response if a “ceiling effect” limits further improvement in that parameter. Similar research showed low exercise doses can effectively reduce cardiovascular disease or cardiovascular risk factor prevalence, but higher exercise doses do not yield additional benefits [31].

Our present work showed an increase in baPWV for the health care mode group and unchanged results for the scientific fitness mode group. These changes differed from those reported in previous studies [32-35]. However, studies included in meta-analyses investigating the effects of exercise on arterial stiffness also report positive evidence [36-38] for aerobic exercise, with patients with isolated systolic hypertension being the exception. Resistance exercise has different effects on arterial stiffness depending on type and intensity, and there seem to be no unfavorable effects on arterial stiffness if the training is of low intensity or performed in a slow concentric manner or in the lower limbs of healthy individuals [37]. Furthermore, combined training has no significant effects on arterial stiffness [36]. In our study, more than half of the participants in with healthcare mode (56/97, 58%) and nearly half of the participants in the scientific fitness mode(36/80, 45%) were diagnosed with hypertension, and those in the health care mode group were given a low-intensity aerobic exercise prescription. This partly explains the lack of improvement in baPWV, as blood pressure is affected by many confounding factors (eg, medication). Another explanation might be age. Aging reduces arterial elasticity and has been suggested to be the main precursor of arterial stiffness in different populations [39,40], with this change being more significant in older adults. Noticeably, in our present work, arterial elasticity became worse in participants engaging in aerobic exercise only, but there was no significant difference in baPWV among community dwellers engaging in both aerobic exercise and resistance training. Given the undifferentiated baseline data, it may be that the potential effectiveness of this intelligent, personalized, structured exercise prescription intervention has been demonstrated. However, additional well-designed randomized controlled trials are needed before any final recommendations can be formulated.

This 12-month intervention study showed no difference before and after intervention in bone mineral density for either group. Although some previous trials have reported the effectiveness of intervention [41], there is currently insufficient evidence to recommend exercise for improving bone mineral density [42]. One meta-analysis reported that the positive results were small, nonsignificant, and with a large and statistically significant amount of heterogeneity. There were also some consistent results detecting no significant changes in breast cancer survivors [43], or overweight and obese older adults [44].

Evidence does consistently suggest that exercise leads to significant improvements in physical fitness, increased flexibility, agility, and strength [32,41,45-48]. However, in our present work, physical fitness showed a decline in both exercise prescription groups. The inclusion of older participants may partly explain this. Reduction in physical fitness, including reduction of muscle strength in both the upper and lower limbs, and changes in flexibility, agility, and endurance, were equal for both men and women and was likely due to the aging process, which was discussed in a previous study [49].

No adverse events were reported for any of the interventions. Although our prescription included step-by-step instructions with pictures and participants could call community health center staff to report any uncomfortable conditions during exercise, there are safety aspects worth considering that should

XSL•FO
RenderX

be more strictly monitored. We should strengthen the monitoring of exercises so that harmful execution of exercises can be quickly noted and addressed.

There were several limitations to this research. First, this study was conducted in 1 geographic location, which limits the generalizability of observations and hinders the ability to identify population differences. Second, we only studied a single group of middle-aged and older adults and did not include a control group, impeding any ability to draw conclusions on initial intervention effectiveness. Third, the adherence to exercise prescription was only based on a self-report of participants. We lacked a objective and scientific means of monitoring the entire program. Fourth, as a feasibility and single-arm study, we failed to control age and disease-related confounding factors, which might have potentially influenced our observations. Future studies using a randomized controlled intervention protocol and employing app-based, wearable devices are encouraged to expand on this effort.

The observations suggested that our exercise prescription intervention program might promote certain health outcomes, such as cardiovascular function and body composition in middle-aged and older adult Chinese community dwellers. However, we are unable to recommend this program because of the existing limitations. Nonetheless, we recommend that older adults with a range of diseases begin exercise under supervised instruction when initiating training. The benefit is clear, and “start to exercise” should be the top priority for all older adults.

Acknowledgments

The authors thank Lan Xie and Lin Liu at the Community Health Service Centre for administrative assistance. This work was supported by the Anhui Provincial Major Science and Technology Project (grant no. 18030801133) and the Science and Technology Service Network Initiative (grant no. KFJ-STS-ZDTP-079). The funding bodies played no role in the design of the study; the collection, analysis, or interpretation of data; or in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Procedure for measurement with main instruments.
[DOCX File , 278 KB - jmir_v23i5e28221_app1.docx]

Multimedia Appendix 2
Examples of 3 modes of exercise prescription.
[DOCX File , 804 KB - jmir_v23i5e28221_app2.docx]

Multimedia Appendix 3
Algorithms for different exercise types.
[DOCX File , 121 KB - jmir_v23i5e28221_app3.docx]

Multimedia Appendix 4
Changes of health outcomes in the scientific fitness mode (n=80).
[DOCX File , 18 KB - jmir_v23i5e28221_app4.docx]

References

1. Thornton JS, Frémont P, Khan K, Poirier P, Fowles J, Wells GD, et al. Physical activity prescription: a critical opportunity to address a modifiable risk factor for the prevention and management of chronic disease: a position statement by the Canadian Academy of Sport and Exercise Medicine. Clin J Sport Med 2016 Jul;26(4):259-265. [doi: 10.1097/JSM.0000000000000363] [Medline: 27359294]
2. Hamlin MJ, Yule E, Elliot CA, Stoner L, Kathiravel Y. Long-term effectiveness of the New Zealand Green Prescription primary health care exercise initiative. Public Health 2016 Nov;140:102-108. [doi: 10.1016/j.puhe.2016.07.014] [Medline: 27569778]
3. Shad BJ, Wallis G, van Loon LJC, Thompson JL. Exercise prescription for the older population: The interactions between physical activity, sedentary time, and adequate nutrition in maintaining musculoskeletal health. Maturitas 2016 Nov;93:78-82. [doi: 10.1016/j.maturitas.2016.05.016] [Medline: 27338978]
4. Luan X, Tian X, Zhang H, Huang R, Li N, Chen P, et al. Exercise as a prescription for patients with various diseases. J Sport Health Sci 2019 Sep;8(5):422-441 [FREE Full text] [doi: 10.1016/j.jshs.2019.04.002] [Medline: 31534817]

XSL•FO
RenderX

5. Wilund KR, Viana JL, Perez LM. A critical review of exercise training in hemodialysis patients: personalized activity prescriptions are needed. Exerc Sport Sci Rev 2020 Jan;48(1):28-39 [FREE Full text] [doi: 10.1249/JES.0000000000000209] [Medline: 31453844]
6. Garnacho-Castaño MV, Domínguez R, Muñoz González A, Feliu-Ruano R, Serra-Payá N, Maté-Muñoz JL. Exercise prescription using the Borg Rating of Perceived Exertion to improve fitness. Int J Sports Med 2018 Feb;39(2):115-123. [doi: 10.1055/s-0043-120761] [Medline: 29190852]
7. Garzon M, Gayda M, Nigam A, Comtois A, Juneau M. Immersible ergocycle prescription as a function of relative exercise intensity. J Sport Health Sci 2017 Jun;6(2):219-224 [FREE Full text] [doi: 10.1016/j.jshs.2015.12.004] [Medline: 30356614]
8. Bray NW, Smart RR, Jakobi JM, Jones GR. Exercise prescription to reverse frailty. Appl Physiol Nutr Metab 2016 Oct;41(10):1112-1116 [FREE Full text] [doi: 10.1139/apnm-2016-0226] [Medline: 27649859]
9. Kraemer RR, Landefeld CS. Exercise programs in older adults-a prescription for fall reduction. JAMA Intern Med 2019 Mar 01;179(3):405-406. [doi: 10.1001/jamainternmed.2018.5389] [Medline: 30592478]
10. Krops LA, Bouma AJ, Van Nassau F, Nauta J, van den Akker-Scheek I, Bossers WJ, et al. Implementing individually tailored prescription of physical activity in routine clinical care: protocol of the Physicians Implement Exercise = Medicine (PIE=M) Development and Implementation Project. JMIR Res Protoc 2020 Nov 02;9(11):e19397 [FREE Full text] [doi: 10.2196/19397] [Medline: 33136060]
11. Chatterjee R, Wolman R. Exercise is medicine: a case study of an exercise prescription clinic in the NHS. Br J Gen Pract 2019 Jun;69(683):307-308 [FREE Full text] [doi: 10.3399/bjgp19X704033] [Medline: 31147338]
12. Porcelli S, Bianchi G, Agnello L, Manferdelli G, Mastropietro A, Pigozzi F. Exercise prescription for health: Italian perspective. Italian guidelines for exercise prescription in healthy adults (18-65 years). Medicina Dello Sport 2020 Mar;73(1):140-164. [doi: 10.23736/s0025-7826.20.03689-3]
13. Tuka V, Linhart A. Personalised exercise prescription: finding the best for our patients. Eur J Prev Cardiol 2020 Sep;27(13):1366-1368. [doi: 10.1177/2047487319884376] [Medline: 31640416]
14. Hansen D, Niebauer J, Cornelissen V, Barna O, Neunhäuserer D, Stettler C, et al. exercise prescription in patients with different combinations of cardiovascular disease risk factors: a consensus statement from the Expert Working Group. Sports Med 2018 Aug;48(8):1781-1797. [doi: 10.1007/s40279-018-0930-4] [Medline: 29729003]
15. Woodward MJ, Lu CW, Levandowski R, Kostis J, Bachmann G. The exercise prescription for enhancing overall health of midlife and older women. Maturitas 2015 Sep;82(1):65-71. [doi: 10.1016/j.maturitas.2015.03.020] [Medline: 25963099]
16. Papadopoulou SK, Papandreou D, Tassoulas E, Biskanaki F, Kalogiannis S, Hassapidou MN. Gender and exercise in relation to obesity in Greek elderly population. Int J Environ Res Public Health 2020 Sep 09;17(18):1-9 [FREE Full text] [doi: 10.3390/ijerph17186575] [Medline: 32917002]
17. Durst J, Roesel I, Sudeck G, Sassenberg K, Krauss I. Effectiveness of human versus computer-based instructions for exercise on physical activity-related health competence in patients with hip osteoarthritis: randomized noninferiority crossover trial. J Med Internet Res 2020 Sep 28;22(9):e18233 [FREE Full text] [doi: 10.2196/18233] [Medline: 32985991]
18. Liu J, Li G. Practical Exercise Prescription. 1st Edition. Harbin: Heilongjiang Science and Technology Press; 1993.
19. Thompson PD, Arena R, Riebe D, Pescatello LS. ACSM's new preparticipation health screening recommendations from ACSM's guidelines for exercise testing and prescription, ninth edition. Current Sports Medicine Reports 2013;12(4):215-217. [doi: 10.1249/jsr.0b013e31829a68cf]
20. Hansen D, Dendale P, Coninx K, Vanhees L, Piepoli MF, Niebauer J, et al. The European Association of Preventive Cardiology Exercise Prescription in Everyday Practice and Rehabilitative Training (EXPERT) tool: A digital training and decision support system for optimized exercise prescription in cardiovascular disease. Concept, definitions and construction methodology. Eur J Prev Cardiol 2017 Jul;24(10):1017-1031. [doi: 10.1177/2047487317702042] [Medline: 28420250]
21. Hansen D, Rovelo Ruiz G, Doherty P, Iliou M, Vromen T, Hinton S, et al. Do clinicians prescribe exercise similarly in patients with different cardiovascular diseases? Findings from the EAPC EXPERT working group survey. Eur J Prev Cardiolog 2018 Feb 28;25(7):682-691. [doi: 10.1177/2047487318760888]
22. Foccardi G, Hansen D, Quinto G, Favero C, Coninx K, Ruiz G, et al. How do general practitioners assess physical activity and prescribe exercise in patients with different cardiovascular diseases? An Italian pilot study. Eur J Prev Cardiol 2020 May 16:616. [doi: 10.1177/2047487320925221] [Medline: 33611460]
23. Husebø AML, Dyrstad SM, Søreide JA, Bru E. Predicting exercise adherence in cancer patients and survivors: a systematic review and meta-analysis of motivational and behavioural factors. J Clin Nurs 2012 Nov 17;22(1-2):4-21. [doi: 10.1111/j.1365-2702.2012.04322.x]
24. Forhan M, Zagorski BM, Marzonlini S, Oh P, Alter DA. Predicting exercise adherence for patients with obesity and diabetes referred to a cardiac rehabilitation and secondary prevention program. Can J Diabetes 2013 Jun;37(3):189-194. [doi: 10.1016/j.jcjd.2013.03.370] [Medline: 24070842]
25. van der Leeden M, Stuiver MM, Huijsmans R, Geleijn E, de Rooij M, Dekker J. Structured clinical reasoning for exercise prescription in patients with comorbidity. Disabil Rehabil 2020 May;42(10):1474-1479. [doi: 10.1080/09638288.2018.1527953] [Medline: 30574815]
26. Yang XY. Study on the design of individualized exercise intervention program based on behavior change theory [dissertation]. Hefei: University of Science and Technology of China; 2019.

XSL•FO
RenderX

27. Magutah K, Thairu K, Patel N. Effect of short moderate intensity exercise bouts on cardiovascular function and maximal oxygen consumption in sedentary older adults. BMJ Open Sport Exerc Med 2020;6(1):e000672 [FREE Full text] [doi: 10.1136/bmjsem-2019-000672] [Medline: 32180993]
28. Wu Q, Zhang X, Xu Y, Wang M, Wang Y, Yang X, et al. A cross-section study of main determinants of arterial stiffness in Hefei area, China. Int Angiol 2019 Apr;38(2):150-156 [FREE Full text] [doi: 10.23736/S0392-9590.19.04078-1] [Medline: 30938496]
29. Huang J, Lai Q, Wang D, Yin H, Liao J, Wang S, et al. Effects of exercise training with dietary restriction on arterial stiffness, central hemodynamic parameters and cardiac autonomic function in obese adolescents. Diabetes Metab Syndr Obes 2019;12:2157-2163. [doi: 10.2147/DMSO.S223514] [Medline: 31802923]
30. Mann TN, Lamberts RP, Lambert MI. High responders and low responders: factors associated with individual variation in response to standardized training. Sports Med 2014 Aug;44(8):1113-1124. [doi: 10.1007/s40279-014-0197-3] [Medline: 24807838]
31. Maessen MF, Verbeek AL, Bakker EA, Thompson PD, Hopman MT, Eijsvogels TM. Lifelong exercise patterns and cardiovascular health. Mayo Clin Proc 2016 Jun;91(6):745-754. [doi: 10.1016/j.mayocp.2016.02.028] [Medline: 27140541]
32. Park W, Jung W, Hong K, Kim Y, Kim S, Park H. Effects of moderate combined resistance- and aerobic-exercise for 12 weeks on body composition, cardiometabolic risk factors, blood pressure, arterial stiffness, and physical functions, among obese older men: a pilot study. Int J Environ Res Public Health 2020 Oct 03;17(19):1-12 [FREE Full text] [doi: 10.3390/ijerph17197233] [Medline: 33022918]
33. Otsuki T, Namatame H, Yoshikawa T, Zempo-Miyaki A. Combined aerobic and low-intensity resistance exercise training increases basal nitric oxide production and decreases arterial stiffness in healthy older adults. J Clin Biochem Nutr 2020 Jan;66(1):62-66 [FREE Full text] [doi: 10.3164/jcbn.19-81] [Medline: 32001958]
34. Shiotsu Y, Watanabe Y, Tujii S, Yanagita M. Effect of exercise order of combined aerobic and resistance training on arterial stiffness in older men. Exp Gerontol 2018 Oct 01;111:27-34. [doi: 10.1016/j.exger.2018.06.020] [Medline: 29953951]
35. Pereira T, Santos F, Cipriano I. Effects of a personalized physical exercise program in the arterial stiffness in older adults. In: ARTRES. 2019 Presented at: 29th European Meeting of Hypertension and Cardiovascular Protection of the European-Society-of-Hypertension; June 6 2019; Milan, Italy. [doi: 10.2991/artres.k.191126.001]
36. Sardeli AV, Gáspari AF, Chacon-Mikahil MP. Acute, short-, and long-term effects of different types of exercise in central arterial stiffness: a systematic review and meta-analysis. J Sports Med Phys Fitness 2018 Jun;58(6):923-932. [doi: 10.23736/S0022-4707.17.07486-2] [Medline: 29877680]
37. Li Y, Hanssen H, Cordes M, Rossmeissl A, Endes S, Schmidt-Trucksäss A. Aerobic, resistance and combined exercise training on arterial stiffness in normotensive and hypertensive adults: A review. Eur J Sport Sci 2015;15(5):443-457. [doi: 10.1080/17461391.2014.955129] [Medline: 25251989]
38. Ashor AW, Lara J, Siervo M, Celis-Morales C, Mathers JC. Effects of exercise modalities on arterial stiffness and wave reflection: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(10):e110034 [FREE Full text] [doi: 10.1371/journal.pone.0110034] [Medline: 25333969]
39. London GM, Pannier B, Safar ME. Arterial stiffness gradient, systemic reflection coefficient, and pulsatile pressure wave transmission in essential hypertension. Hypertension 2019 Dec;74(6):1366-1372. [doi: 10.1161/HYPERTENSIONAHA.119.13387] [Medline: 31679422]
40. Mynard JP, Clarke MM. Arterial stiffness, exercise capacity and cardiovascular risk. Heart Lung Circ 2019 Nov;28(11):1609-1611. [doi: 10.1016/j.hlc.2019.09.002] [Medline: 31627826]
41. Marques EA, Mota J, Viana JL, Tuna D, Figueiredo P, Guimarães JT, et al. Response of bone mineral density, inflammatory cytokines, and biochemical bone markers to a 32-week combined loading exercise programme in older men and women. Arch Gerontol Geriatr 2013;57(2):226-233. [doi: 10.1016/j.archger.2013.03.014] [Medline: 23623588]
42. Kelley G, Kelley K, Kohrt W. Corrigendum to “Exercise and bone mineral density in men: a meta-analysis of randomized controlled trials” [Bone 53 (2013) 103–111]. Bone 2013 Sep;56(1):30. [doi: 10.1016/j.bone.2013.05.007]
43. Boing L, Vieira MDCS, Moratelli J, Bergmann A, Guimarães ACDA. Effects of exercise on physical outcomes of breast cancer survivors receiving hormone therapy - A systematic review and meta-analysis. Maturitas 2020 Nov;141:71-81. [doi: 10.1016/j.maturitas.2020.06.022] [Medline: 33036706]
44. Beavers KM, Beavers DP, Martin SB, Marsh AP, Lyles MF, Lenchik L, et al. Change in bone mineral density during weight loss with resistance versus aerobic exercise training in older adults. J Gerontol A Biol Sci Med Sci 2017 Oct 12;72(11):1582-1585 [FREE Full text] [doi: 10.1093/gerona/glx048] [Medline: 28379325]
45. Dib MM, Tomeleri CM, Nunes JP, Cunha PM, Ribeiro AS, Schiavoni D, et al. Effects of three resistance exercise orders on muscular function and body composition in older women. Int J Sports Med 2020 Dec;41(14):1024-1031. [doi: 10.1055/a-1192-5205] [Medline: 32688414]
46. Park S, Kwak Y, Pekas EJ. Impacts of aquatic walking on arterial stiffness, exercise tolerance, and physical function in patients with peripheral artery disease: a randomized clinical trial. J Appl Physiol (1985) 2019 Oct 01;127(4):940-949 [FREE Full text] [doi: 10.1152/japplphysiol.00209.2019] [Medline: 31369328]
47. Dieli-Conwright CM, Courneya KS, Demark-Wahnefried W, Sami N, Lee K, Sweeney FC, et al. Aerobic and resistance exercise improves physical fitness, bone health, and quality of life in overweight and obese breast cancer survivors: a

XSL•FO
RenderX

randomized controlled trial. Breast Cancer Res 2018 Oct 19;20(1):1-10 [FREE Full text] [doi: 10.1186/s13058-018-1051-6] [Medline: 30340503]
48. Choi M, Sohng K. The effects of floor-seated exercise program on physical fitness, depression, and sleep in older adults: a cluster randomized controlled trial. International Journal of Gerontology 2018 Jun;12(2):116-121. [doi: 10.1016/j.ijge.2017.06.003]
49. Milanović Z, Pantelić S, Trajković N, Sporiš G, Kostić R, James N. Age-related decrease in physical activity and functional fitness among elderly men and women. Clin Interv Aging 2013;8:549-556 [FREE Full text] [doi: 10.2147/CIA.S44112] [Medline: 23723694]

Abbreviations

ABI: ankle/brachial systolic pressure index
ACSM: American College of Sports Medicine
baPWV: brachial-ankle pulse wave velocity
DBP: diastolic blood pressure
IPEPP: intelligent personalized exercise prescription program
MD: mean difference
NCD: noncommunicable chronic disease
SBP: systolic blood pressure
SEVR: subendocardial viability ratio
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.02.21; peer-reviewed by X Yang, D Hansen; comments to author 22.03.21; revised version received 15.04.21; accepted 29.04.21; published 24.05.21.

Please cite as:
Sun T, Xu Y, Xie H, Ma Z, Wang Y
Intelligent Personalized Exercise Prescription Based on an eHealth Promotion System to Improve Health Outcomes of Middle-Aged and Older Adult Community Dwellers: Pretest–Posttest Study
J Med Internet Res 2021;23(5):e28221
URL: https://www.jmir.org/2021/5/e28221
doi:10.2196/28221
PMID:34028359

©Ting Sun, Yang Xu, Hui Xie, Zuchang Ma, Yu Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.05.2021. 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 Social Media Organizational Productivity Model: Insights From Public Health Professionals

Hamad Ghaleb Dailah Sr[1], PhD; Muhammad Naeem[2], DBA

[1]Faculty of Nursing, Jazan University, Jazan, Saudi Arabia

[2]University of Gloucestershire, Cheltenham, United Kingdom

Corresponding Author:
Hamad Ghaleb Dailah Sr, PhD
Faculty of Nursing
Jazan University
Al Maarefah Rd
Jazan
Saudi Arabia
Phone: 966 17 329 5000
Email: drhamaddilah@gmail.com

Abstract

Background: Many previous studies have explored socialization-oriented social media (SM), but their reach has been limited to the context of information exchange for common personal interests. This study focuses on work-oriented SM, which can enhance organizational networking and productivity levels in the context of public hospitals.

Objective: This study aims to provide a theoretical framework to explain how the use of SM can enhance the skills of health professionals and levels of organizational productivity in uncertain environments.

Methods: A total of 2 distinct forms of data collection techniques were combined: focus groups and semistructured interviews. Both were conducted with doctors and nurses in Saudi public sector hospitals.

Results: The findings reveal that the use of SM can create professional socialization at the level of the institution, and this can enhance skills, knowledge, decision making, and the overall level of organizational productivity. The increasing use of SM creates collaboration between health experts (particularly endocrinologists and pulmonologists in this case) who arrange video calls to share best practices in terms of medication, diet, and health care plans for patients with multiple diseases. Many of these patients are particularly vulnerable, given the wider context of the current global pandemic.

Conclusions: This study culminates in the Social Media Organizational Productivity model, which provides insights into how SM has increased the accessibility of health professionals through the use of technology. Access to such professionals creates a patient-centric approach and a culture of shared communication for dealing with high-risk patients during the current global pandemic.

(J Med Internet Res 2021;23(5):e23792) doi:10.2196/23792

KEYWORDS

social media; professional socialization; uncertainty; institutional theory; motivation; public hospital; health professionals

Introduction

Background

The flexible, agile, collaborative, informal, unstructured, and spontaneous nature of social media (SM) platforms has given rise to more and better social interactions, social support, knowledge sharing, and social expectations [1,2]. There are various kinds of SM, such as personal platforms (ie, Skype and WhatsApp) and organizational SM platforms (ie, Microsoft Yammer). Thus, SM is heterogeneous at both the personal and organizational levels. However, contemporary employees cannot differentiate between personal and corporate SM. On the basis of this heterogeneity and motivated by the influence of the SM world, workplaces have begun to embrace SM for socialization and work-oriented purposes [3,4]. Work-oriented SM types include web-based platforms that can be used in the workplace to facilitate collaboration. These forms of SM can also generate resources and facilitate the sharing of work-related content and information. They can support internal organizational communication and can be used to track events and

XSL•FO
RenderX

workstreams. They are also reliable forms of support for task management [5,6]. However, socialization-oriented SM includes web-based platforms that facilitate the sharing of personal and social information and that enable people to develop expressive links that can shape identities through normative expectations and emotional and social forms of support [6,7].

Existing literature on SM has tended to focus on exploring socialization-oriented SM, which is limited to information exchange for personal common interests. However, this study focuses on work-oriented SM that can influence organizational productivity in the setting of Saudi public sector hospitals. SM plays a very important role in supporting communication among employees in organizations [8]. A key previous study highlighted the importance of relationships and social interactions to achieve common objectives [8]. Virtuality theory advocates the use of SM if it is required to alter patterns of social exchange ties. However, it overlooks the consequences of such social exchange ties in organizational settings. Previous studies have measured the impact of patient tweets on hospital quality improvements, but these studies failed to produce the expected results [6,9,10]. Approximately 60% of physicians use SM to monitor patient health and education. Furthermore, these physicians use SM sites to inform patient drug adherence and behavioral changes that can enhance the level of patient compliance and positive outcomes [6,9,10]. Conversely, a survey of 480 physicians revealed that 68% of physicians hesitate to use SM sites for patient, personal, and professional connections. However, there is little research on how the use of SM for communication purposes influences organizational productivity in the Saudi health sector setting. High-quality SM relationships make a significant contribution to employee performance, but there is little understanding of this contribution in public hospitals in Saudi Arabia [11].

There is evidence of the popularity of SM among Saudi men and women, as approximately 41% of the Saudi population actively uses various SM platforms [2]. According to recent statistics, some 9.9 million Saudi Arabians regularly use Twitter. Saudi Arabians are the fourth highest users of Twitter in the world [12]. Facebook, WhatsApp, and Twitter are used to exchange audio, videos, text, views, and information and to meet personal and organizational interests. The Saudi government has restricted the use of some SM platforms because of social campaigns by SM activists between 2013 and 2017 [13]. However, after this time, the Saudi government lifted the ban and allowed the use of the SM platforms [13]. Public hospitals in Saudi Arabia are under the control of the Saudi Ministry of Health. It is therefore interesting to analyze the specific rules, regulations, policies, hierarchal structure, and power uses that can influence the use of SM platforms, even for productive purposes in Saudi public sector hospitals.

Although an ample body of literature has explored the effective use of SM at individual levels as well as in the organizational communication context in Western culture [14,15], there is evidence that Arab public organizational culture is unique and different from Western public organizational culture. This might be because of some inherent cultural attributes, such as authoritative management styles, hierarchal structures, a preference for top-down communication, low staff empowerment, and strong social networking among those with sufficient resources. Work overload and favoritism might also be factors. These are some of the unique features of Saudi public sector organizations that strongly influence the use of SM platforms at both the individual and organizational levels. There is little understanding of how public hospitals in Arab countries such as Saudi Arabia manage SM use. This study introduces a theoretical framework to guide how SM can drive productivity in Saudi public sector hospitals.

Literature Review

Deep penetration of SM has been observed in many workplaces. Nowadays, there is an increasing trend among companies to strategically implement such tools to improve their organizational activities and support their workforces [16]. Leading organizations regularly use personal, public, and popular SM such as LinkedIn, Twitter, and Facebook to improve employee engagement, customer service, talent recruitment, marketing, and knowledge sharing [17,18]. Research has described SM use in health information management as “the activities that people perform to acquire, organize, maintain, share, retrieve, and use health information items to complete healthcare tasks and fulfil their needs” [19]. The increasing use of SM in the workplace has led to the emergence of various socially connected software that companies commonly use to pursue their business activities and goals [5]. In addition, where the aim is to support employees, such use is considered promising for excellent organizational performance [9,20].

In addition to extending access to public and general SM platforms such as Twitter and Facebook in workplaces [10,20], more emphasis is being placed on developing and implementing professional and more specific SM technologies. Examples include IBM connections, Slack, Jive, Facebook Workplace, DingTalk, and Microsoft Yammer. Consequently, it has become possible for different types of SM platforms to coexist in organizations. These can be used by workers for work and/or during work time for information exchange and effectiveness in workplaces. However, companies and executives may sometimes perceive SM as controversial [16]. This begs the question: How does the utilization of different kinds of SM platforms in the workplace influence organizational productivity, and what can be learned specifically from public sector hospitals in Saudi Arabia? Ability, motivation, and opportunity (AMO) theory holds that if employees are competent, skilled, and motivated, then they can create more opportunities at the personal and organizational levels. Therefore, organizations should provide training and career advancement opportunities to employees, as this can enhance their skills and motivation and can drive organizational productivity.

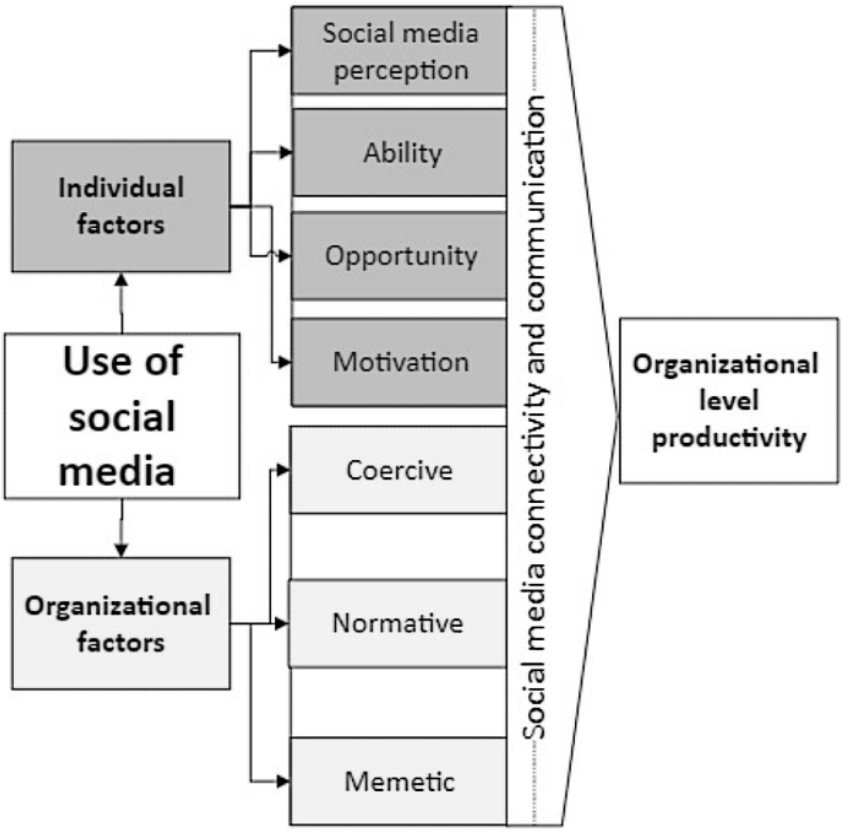
There are various fields in which institutional theory has been applied, such as political science, economics, organizational theory, and sociology. However, there is little understanding about the application of institutional theory in public hospital settings. In the context of institutional theory, organizations embed themselves within different environments in which they may acquire legitimacy from imitative, coercive, or other kinds of institutional forces. Moreover, institutional theory implies that corporate decisions are partially driven by rational

XSL•FO
RenderX

effectiveness goals and partially by legitimacy concerns and cultural and social factors. Companies adhere to different institutional values or norms to acquire a *fit* and/or to integrate into the surrounding environment. Isomorphism provides an in-depth understanding of the institutional environment. By incorporating institutional rules within their own structures, organizations become more homogeneous and similar in structure over time. This is particularly the case within particular institutional environments and contexts, such as how public hospitals in Arab countries, such as Saudi Arabia, manage SM use.

The 3 common methods of isomorphism identified by DiMaggio and Powell are coercive isomorphism (firms are either compelled to adhere to rules or adopt to structures), normative isomorphism (firms adopt forms as professionals within the firm who claim their superiority over others), and mimetic isomorphism (one firm copies another firm, often because of uncertainty). Institutionalization is a process that involves modifying behaviors to fulfill social expectations that are derived from different norms such as regulatory structures, professions, laws, cultural practices, and government agencies. All these actors exert pressure on firms to operate in a particular manner. Institutional theory is considered a powerful and popular explanatory tool for any analysis of organizational behavior that seeks to address different forms of organizational change. In contrast, the concept upon which most classical approaches are based is that organizations are dominated mainly by role, personal preferences, and the interest of individuals, as well as by rational actors. Likewise, new institutionalism primarily focuses on the formative role of organizations. In this sense, the core postulate is that rational actors within organizations always look out for their personal interests and exploit particular institutional limitations in pursuit of these. This view suggests that patterns of organizational actions are not shaped by instrumental calculations, but by institutional forces, such as cultural scripts and norms. The deep penetration of organizations in social environments could expedite or compromise their use of SM platforms for organizational productivity purposes (Figure 1). This suggests that organizational procedures and structures often reflect the environmental expectations that are required to enhance the use of SM to make the organization more successful. Institutional theory is based on the idea that organizations always want to respond to their environmental demands with the purpose of leveraging competitive advantage.

Figure 1. Theoretical framework based on institutional and ability, motivation, and opportunity theories.



Methods

Overview

There is an increasing trend among researchers to use social constructivism to understand and synthesize the different realities of multiple disciplines, including the health sciences, philosophy, sociology, and psychology [1,21,22]. In particular, health science researchers extensively use social constructivism to gain in-depth insights into how health care professionals obtain and integrate new knowledge. They also tend to explore the kinds of behaviors they can develop to improve their own decision-making performance [22]. According to social constructivists, the social world in which we live has unique

XSL•FO
RenderX

local, cultural, and social contexts, and all these contexts vary from location to location, from individual to individual, and from nation to nation [17,23]. Unique and different knowledge insights are, thus, produced through the lens of social constructivism because they critique shared beliefs, shared meaning, shared language, collective cultures, values, and social norms [21,23]. Social constructivists construct knowledge on the basis of cultural and social values. Therefore, they prefer to understand, see, and shape things in accordance with the viewpoints of social actors. The researchers were highly involved in this study to capture and better understand the issues under investigation. Institutional forces, national cultures, individual capabilities and abilities, social structures, organizational policies, the sharing and exchanging of knowledge, and SM are all subjective realities. Each of these influences the organizational productivity of public hospitals. Through the research process, this study clearly shows how social constructivism, analytical techniques, and data collection are interconnected.

There is limited literature available that has sought to uncover the influence of SM on employee work productivity in a public sector hospital. Therefore, this study is exploratory in nature because it aims to capture rich insights using social constructivism and interpretive methods. Social constructivists believe that there are multiple environmental and biological factors that have different impacts on individuals [17,21]. Interpretive methods are considered important where the researcher seeks to understand the social and *real* context of networking technologies and how social networking technologies influence organizational productivity in health care workplaces. Interpretive methods are thus recognized as important, especially when people are interested in knowing how the use of technology influences various institutional, government, cultural, and social structural factors [24,25]. Indeed, interpretive studies are more focused on understanding social contexts (ie, social relationships, peer influence, the division of labor, and the structure of an organization) that are impacted by the actions of individuals [17,21]. Previous studies have indicated that interpretive methods can capture subjective realties, such as individual levels of information and levels of knowledge, expectation, experience, and understanding [17,23]. Other studies have suggested that interpretive or qualitative methods are more useful for capturing the exploratory elements of subjects with the support of single or multiple theories [22]. These methods are usually useful for understanding individual perceptions about and sense-making in particular situations.

Sampling and Population

This study primarily focuses on Saudi public sector (government) hospitals, as they are strongly influenced by institutional, cultural, and social forces. Multiple hospitals were selected to gain insights from health care professionals in a variety of government hospitals in Jazan, Saudi Arabia. The hospitals selected for this study included Jizan Hospital, Prince Mohammed Bin Abdul Aziz Hospital, and King Saud Medical City. All these public sector hospitals have a rich technological and institutional history, and many health professionals want to work in them to gain valuable experience.

As many of the doctors in these hospitals are visiting professionals, it was not possible to fully describe the sampling framework. Moreover, although 2 of the hospitals were willing to provide a sampling framework, the others declined to provide any sampling framework. To address this issue, we decided to use a nonprobability sampling method. In health sciences, purposive sampling is common and is particularly useful when researchers are aware of the sampling framework and when they want to choose informants who have rich knowledge about the study objectives. Therefore, purposive sampling was used to collect data for this study. The 30 participants selected for data collection included nurses and head nurses working in government hospitals. Moreover, the 20 participants selected for focus group interviews included emergency doctors, surgeons, and hospitalists. Multimedia Appendices 1 and 2 indicate the demographic attributes of the selected informants. Their selection was based on the following inclusion criteria:

- Working within public sector hospitals for at least two years
- Older than 18 years
- Willing to voluntarily provide research data
- Had at least one active account on SM

Data Collection Technique

To collect qualitative data, 2 distinct forms of data collection techniques were used: focus group interviews and semistructured interviews. Data collection took about 7 months, and 3 different methods were used to gain in-depth insights into the expectations and experiences of professionals who use SM in the workplace. Semistructured interviews were undertaken because this type of method ensures the maximum participation of respondents throughout the research process. Moreover, “researchers have realized [for] quite some time that researchers are not invisible neutral entities; rather they are part of the interactions they seek to study and influence as well as observe those interactions.” Of real interest here were the experiences of nurses and their perceptions of the different realities of organizational productivity and SM use. Doctors were seen as key sources for verifying and conceptualizing these shared realities.

Interpretative methods are useful for gaining rich insights through observations, semistructured interviews, stories, diaries, and narratives [24,25]. Semistructured interviews are one of the most commonly used interpretative methods among interpretivists because they can elicit rich insights and generate answers to multiple questions such as *what, how, and why* [22,26]. Moreover, the flexibility of semistructured interviews means that it is possible to add or modify the interview questions. Contrary to structured or open-ended interviews, semistructured interviews focused on obtaining particular answers [24,25]. Semistructured interviews were conducted in a face-to-face setting. The following questions were asked during semistructured interviews: How do you see the role of SM in relation to organizational productivity in the workplace? What socialization factors appear useful for increasing your productivity? What are your organizational and individual perceptions of SM use within hospitals? What challenges negatively influence SM use within this hospital? What are the specific public sector policies or factors that influence the use of SM in public sector hospital settings?

XSL•FO
RenderX

Focus group interviews are widely used by researchers interested in accessing the opinions and life stories of individuals in specific situations. Many researchers recommend conducting focus group interviews whenever there is a need to handle discourse in which the image of the research participants is continuously regenerated. Health professionals were invited to participate in focus group interviews. The sample comprised 20 doctors and 30 nurses (Multimedia Appendices 1 and 2). This sample resonates with previous qualitative studies wherein the data saturation point was reached before the 30th interview [24,25]. To maintain social distancing, the social networks of health professionals and the researcher played an important role in identifying respondents with knowledge of SM technologies. Initially, the participants were asked if they would like to volunteer to participate in the interviews. In addition, some health professionals were contacted via SM (such as Twitter, Facebook, and WhatsApp). The researchers sent reminder emails to contact any participants who had not yet confirmed their interview schedule either by Skype or telephone. Interviews through Skype and telephone took place to maintain social distancing during the COVID-19 pandemic.

Interviews with nurses and head nurses lasted between 30 and 60 minutes. Skype videoconferencing calls were initiated to conduct focus group interviews with the selected health professionals. Participants preferred to take calls in their private offices in the hospitals in which they worked. Interview questions were developed based on the expectations and experiences of health professionals in relation to technology. This is an efficient way through which episodic memory can be accessed and complementary insights can be induced [22,26]. During interviews, episodic memory in individuals' neurocognitive memory systems can be activated by asking them about specific and ordered occasions [22,26], such as employee performance and SM, in this study. Moreover, the researcher used the past tense while developing interview questions to ensure the collection of complete data. Technical questions were also used in this study, whenever required. Many studies have recommended using tactical questions, particularly when managing participants seems quite difficult [22,26]. Indeed, tactical questions represent the summary of questions that were initially asked and are also very useful when it comes to addressing the risks identified earlier.

By asking different forms of interview questions, the researcher can optimize conversations with health professionals on the topic of internal generalizability [22]. Moreover, in the context of social constructivism, internal generalizability was possible with the help of data gathered from different respondents in terms of gender, occupation, and age [22]. The sample drawn for this study contained 30 female respondents of different ages (19-50 years) and experiences (Multimedia Appendix 1). The sample contained 20 doctors (Multimedia Appendix 2). Regardless of the sample composition (more males and fewer female participants), no screening questions preceded the interviews to avoid any intentional demographical influence on an otherwise diverse sample. The respondents were also provided with verbatim transcripts of the interviews to enable cross-examination, increase data validity, and improve internal generalizability.

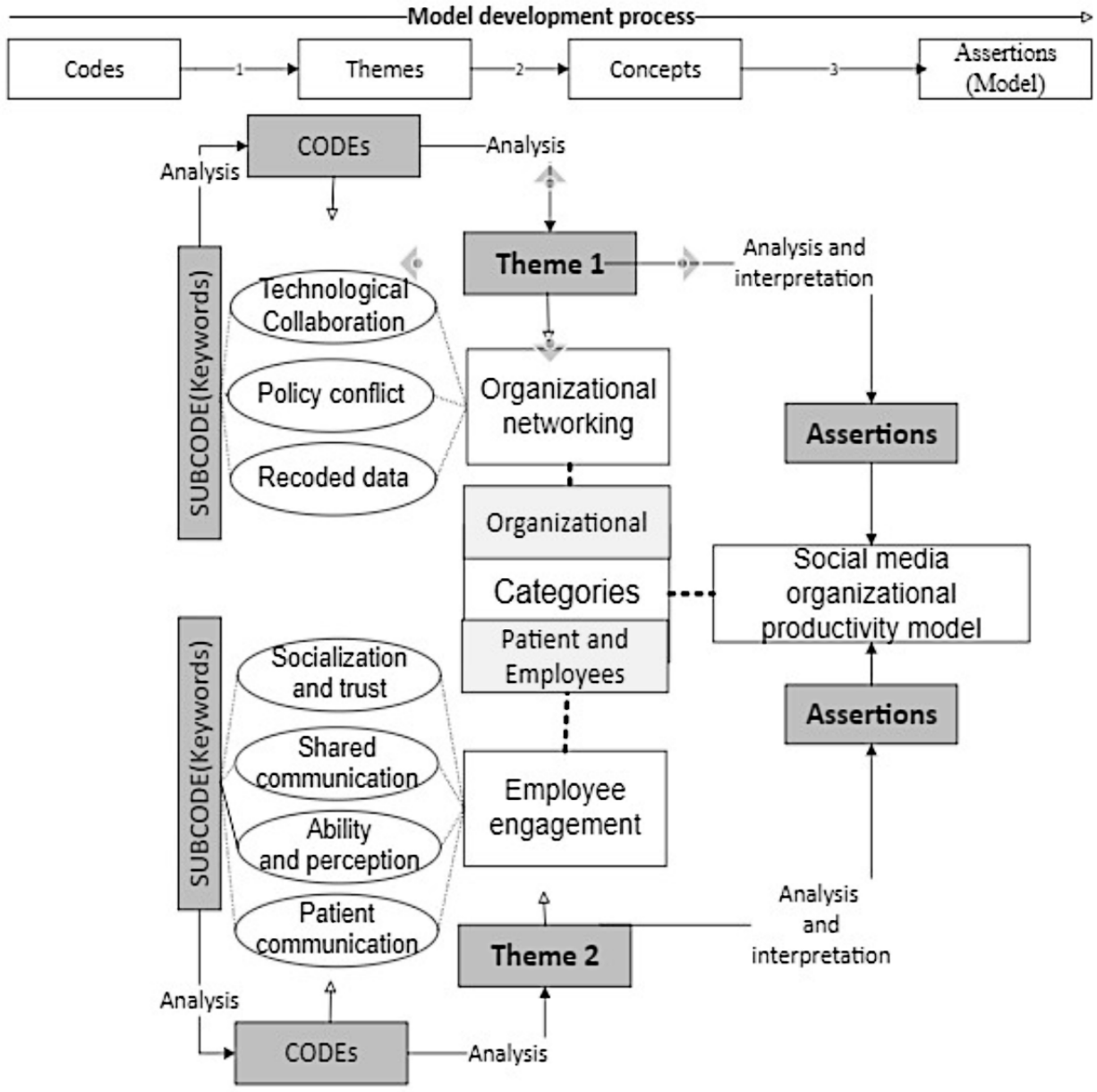
Thematic Analysis

This study used thematic analysis to explore data and to develop a common set of themes or patterns. The researcher also used transcripts of the interviews to develop different initial codes to support the main themes of organizational networking and patient engagement. A total of 3 codes emerged around organizational networking: technological collaboration, policy conflict, and recorded data. In total, 4 codes of employee engagement emerged: socialization and trust, shared communication, ability and perception, and patient communication. After the construction of the main themes, keywords, and codes, the researchers reviewed them to increase both external and internal homogeneity [24,25]. Figure 2 illustrates the thematic analytical procedure of the Social Media Organizational Productivity (SMOP) model.

Data were organized into a verbatim transcript spanning 218 pages. The data were analyzed across 3 key phases. The responses of doctors, head nurses, and nurses were analyzed in the first phase of this process. Next, words that were repeatedly used throughout the interviews were identified. To group these repeated words into codes, the researcher visited them iteratively. Subsequently, the researcher assigned themes to these codes based on their meanings. Further analysis of responses was performed in the second phase. As a result, SM social practices are described. The integration of the experiences of doctors and head nurses was carried out in the third phase, so that a holistic outlook of organizational productivity and SM as a joint experience could be disclosed.

XSL•FO
RenderX

Figure 2. Thematic analysis process.



Trustworthiness of Qualitative Findings

Qualitative studies use confirmability, credibility, dependability, and transferability to maintain rigor and trustworthiness, whereas quantitative studies use reliability, objectivity, and validity to ensure rigor and trustworthiness. *Internal validity should be replaced by credibility, external validity by transferability, reliability by dependability, and objectivity by confirmability.* Although conducting qualitative studies, it is assumed that no *truth* is absolute in nature, but rather multiple realities are accessed. Consequently, generalizing the qualitative findings is not possible. However, credibility, confirmability, transferability, and dependability are, however, very useful measures of rigor and research validity for qualitative studies. Moreover, qualitative studies can improve the transferability of their findings by providing rich descriptions of the research process. This study, for example, produces transferable findings by highlighting the research process and providing a detailed justification for every qualitative data gathering method used. The extent to which the researchers or participants in qualitative studies can evaluate both the recommendations and interpretation of study findings is termed dependability. In this study, the researchers evaluated the data many times to ensure the rigor and trustworthiness of interpretations. Semistructured interviews were conducted to ensure the confirmability of the findings. The use of different qualitative data collection techniques with different respondents increased the credibility of the qualitative findings. In this study, the researcher ensured the credibility of the findings by using focus groups and semistructured interviews with a variety of participants.

XSL•FO
RenderX

Results

Main Theme 1: Organizational Networking

Organizational networking involves exchanging, governing, and organizing sociotechnical networks that can enhance information exchange and social relationships within a formal organization [27]. This study of organizational networking provides ideas about how knowledge, information, communication, and decision making can be carried out in a formal setting. For example, Saudi public sector hospitals operate under the authority of the Ministry of Health. Therefore, these authorities can influence the level of knowledge, information, communication, and decision making, which ultimately influence the level of organizational productivity. To understand organizational networking in Saudi public sector hospitals, this study summarizes the findings into 3 codes: technological collaboration, policy conflict, and recorded data.

Code 1: Technological Collaboration

Technology collaboration refers to the systems and tools that help stakeholders to coordinate and complete tasks effectively, whether working remotely or in an office environment. It is found that nurses use SM instant messaging and video calls to increase interpersonal collaboration. As a result, they can enhance their knowledge, influence the speed at which they work, and increase their organizational productivity. For example, interviewee 19 noted:

> *social media increased peer to peer support as I can consult with my senior nurses through video call and messages and their shared knowledge is always helpful as now, I can be able to complete multiple tasks in limited time.*

There is evidence that SM increases real-time communication, as video calls help health professionals manage patients in intensive care units (ICUs). Specialist doctors cannot be available at all times to deal with critical patients in ICUs. Therefore, SM provides them with the opportunity to manage situations through nurses so that they can assess the condition of multiple patients. For example, interviewee 22 noted:

> *I am working in ICU and many times I faced severe situation as I have to deal with emergency patients who have multiple diseases so we need sometimes urgent video call to different experts so they can advise what I should do until they physically visit to ICU.*

There is evidence that SM has increased the virtual connectivity of different health professionals (ie, endocrinologists and pulmonologists), which means they are in a position to join group discussions with vulnerable patients. This could drive the need for emergency visits to assess vulnerable patients. For example, focus group interviewee 19 stated:

> *Social media helped to virtually connected with my patients as sometimes patients are more critical such as old (unable to visit hospitals) as well as having both diabetes and asthma at the same time so different medical experts (endocrinologist and pulmonologist) can now arrange video call for proper guidance of patients.*

The COVID-19 pandemic has increased many risks to health professionals, patients, and communities. However, SM platforms such as WhatsApp and Skype help users to control situations, especially when there are limited health resources. Therefore, these SM platforms contribute to the care of COVID-19 patients. For example, focus group interviewee 5 noted:

> *Dealing with vulnerable patients has become more complicated, especially when they have the virus, so using Skype and WhatsApp video calls means we can coordinate with each other in a safe community.*

Code 2: Policy Conflict

Policy conflict occurs when there is no professional code of conduct regarding the use of SM platforms. As a result, these social platforms cannot deliver maximum benefits at the organizational level. For example, interviewee 5 noted:

> *The Saudi ministry of health did not develop any organizational policy to use social media in workplace as a result we have not top management support otherwise we can practice it effectively to enhance the speed and collaborative care for patients.*

There is evidence that although the use of SM helps professionals to collaborate and coordinate activities, there is no Saudi government policy to positively use these platforms in the workplace. Therefore, some health professionals use these social platforms, whereas others do not. For example, focus group interviewee 9 said:

> *I am working in public hospitals since 26 years but I experienced when we started to use social media then it saved our time, energy, financial resources, and use of physical infrastructure.*

Some nurses felt that SM platforms really helped them to educate their patients so that they could control emergency situations and find the best way to deliver medicine and diet plans. Furthermore, many used SM to share relevant health care materials with patients, even in the absence of organizational support for the use of SM in public hospital settings. For example, interviewer 18 noted:

> *It is our duty to educate patient so that they can self-manage their diseases and reduce work overload and utilization of human and financial resources. Social media provided us this opportunity to share health care material with patients but there is no organizational policy and leadership support for using these best interactive communication platforms.*

Code 3: Recorded Data

Recorded data refer to SM channels such as YouTube providing facilities to health professionals to record, store, and share data to save time, energy, and resources. In the past, health professionals have tended to verbally communicate to organize tasks that have no official record. As a result, many did not later accept responsibility if anything serious happened to the patients that were being discussed. For example, interviewer 2 noted:

XSL•FO
RenderX

> *Social media helps to generate evidence of communication as, in the past verbal communication is not recorded and sometimes our seniors decline to take responsibility for their recommendations/actions.*

Participants felt more engaged and connected with vulnerable patients and were able to manage emergency situations. As a result, they were able to deal with high numbers of patients, even when they were not at work. For example, focus group interviewee 19 highlighted:

> *I am a diabetes expert and sometimes I receive calls about my diabetes patients and how to clean wounds and stop bleeding in emergency situations...so social media really helps (me) to share YouTube videos of experts, so that these patients can self-manage in emergency situations.*

Recording exchanges on SM also helped professionals to support parents and community education stakeholders by sharing knowledge about, for example, the care of infants among new mothers. For example, focus group interviewee 17 said:

> *I am working in the maternity department so we always share videos to educate the first-time parents about how they should lift their babies...the best positions for burping, and how they should use infant car seats to effectively care for their new-borns.*

Main Theme 2: Employee Engagement

Employee engagement is an increasingly well-researched concept, and many have sought to understand this phenomenon through qualitative and quantitative methods. However, there is less understanding of how engaged health professionals and patients can influence organizational-level productivity, especially in the context of Saudi public sector hospitals. Employee engagement can be explained as motivated and enthusiastic employees involved in positive actions to improve health care services, reputations, and levels of organizational productivity. Although previous studies have provided an understanding of employee engagement from motivational, emotional, and cognitive perspectives, this study examines how SM can increase engagement among health professionals by building socialization and trust and by relying on shared communication, abilities and perceptions, and patient communication.

Code 1: Socialization and Trust

Socialization refers to the enhancement of information exchange to support learning and to improve competencies that can enhance productivity in hospitals. Socialization can promote effective communication among health professionals, and it is therefore useful for building trust in the workplace. For example, interviewee 23 noted:

> *WeChat, Facebook, WhatsApp, and twitter are common social media platforms that we use to communicate and learn from doctors and colleagues. Therefore, we are not more effective and productive compared to the past.*

Socialization promotes the use of social networking platforms (ie, Twitter, WhatsApp, Facebook, and WeChat) for effective communication and is, therefore, an informal, popular, and optimal way of maintaining professional trust in intensive health care settings. The analysis suggests that doctors and nurses are now more connected and engaged and that SM is useful for celebrating individual and organizational achievements. These professionals are now more familiar with SM and can use it in emergency situations. For example, focus group interviewee 13 noted:

> *Nurses and doctors of a particular department are connected through instant messaging apps for celebration of new appointments, promotions, personal educational achievements, parties, and helping others during excessive work overload.*

Some argue against the use of social networking platforms and believe that these have reduced productivity. For example, interviewee 18 noted:

> *I noticed many of my colleagues and seniors even using social media in ICU and operation theatres when they have to focus on their work, so I believe its wastage of time and reduce the work productivity.*

Code 2: Shared Communication

Shared communication refers to explaining, questioning, listening, understanding, acting, advocating, and negotiating over tasks, roles, and responsibilities to enhance organizational productivity. For example, interviewee 13 said:

> *During this global pandemic, both nurse and doctors are extremely busy in ICU and critical operations, but the use of social media increase their interaction, coordination and collaboration especially when health resources are become limited due to increasing the number of patients.*

It is evident that the use of SM has driven coordination and collaboration among doctors and nurses, and they are able to manage more tasks within a limited time frame. The findings also suggest that mental and behavioral issues are heightened among children and older people. Therefore, using SM, doctors and nurses can more effectively discuss health care plans with patients during a global pandemic. For example, focus group interviewee 3 noted:

> *The global pandemic has increased the issues of children and old people who have behaviour and psychological issues. So, the use of social media facilitates shared communication so parents and other people may know what care should be more effective without visiting in hospitals.*

The findings suggest that multiple doctors now collaborate through SM to discuss health care plans with patients who have various diseases and who are vulnerable during global pandemics. For example, focus group interviewee 4 said:

> *I have some patients who have type 2 diabetes, cancer, asthma issues and they are older therefore it is not risk free to meet with them physically. Therefore, sometimes multiple doctors (Trauma surgeon, Endocrinologist, and Pulmonologist)*

XSL•FO
RenderX

conduct conference call for discussing the medication and health care plan.

Code 3: Ability and Perception

Ability refers to the particular skills that we elicit from sharing practical knowledge and experiences through SM platforms. The interviewees had mixed views, and some acknowledged how groups on SM updated their skills and performance, whereas others complained that SM platforms increased expectations when it came to response times. For example, interviewee 9 noted:

> *Our hospital IT department is not very skilled, so we got lower motivational and technical support. Some of our colleagues also have perception that use of these social media platforms created more stress and pressure from our seniors who always demand quick response and action.*

However, focus group interviewee 14 said:

> *Our profession is required to continuously learn and improve our ability, so I have joined many YouTube channels and WhatsApp groups where we are getting opportunities to learn, knowledge exchange, and improve our performance.*

SM has, therefore, connected professionals who have relevant, rich experiences and practical knowledge of the different workplaces at national and international levels. Therefore, there are opportunities for junior staff to increase their skills by learning through experience. For example, interviewee 14 noted:

> *We (head nurses of different hospitals and countries) are now connected with WhatsApp and Facebook groups where we sharing content related to advance healthcare techniques, so we experienced that our level of performance is now more improved.*

Code 4: Patient Communication

Keywords: patient communication, health care, high risks, education, awareness, skills, knowledge.

SM helps to create better health services and patient-centered care. Therefore, doctors are more engaged, particularly with patients who require immediate help to recover from critical situations (ie, pneumonia, chest infections, and asthma). For example, focus group interviewee 9 noted:

> *Instant messaging apps have created direct communication (audio, video, text messaging) between doctors and patients. So now I can more effectively work on asthma patients regarding diet plan, smoking/alcohol habit, inhaler use, and how to control stress in emergency situations.*

In the presence of COVID-19, both public and private hospitals have insufficient resources to provide health care to every mother. Therefore, doctors often only engage with mothers who are struggling because of high-risk pregnancies. For instance, focus group interviewee 6 highlighted:

> *Due to the spread of Covid-19, we cannot ask every patient to come in hospital therefore those mothers who have high risks birth and pregnancies may only contact us through social media and telephone when they need caesareans and interventions.*

The findings highlight that the global pandemic has increased the responsibilities of nurses and doctors who are now less able to provide high standard health care services. However, SM platforms provide them with an opportunity to disseminate educational materials about chronic diseases so that emergency visits can be reduced. For example, focus group interviewee 5 noted:

> *Every asthma patient has different level of complications, education, awareness, and knowledge so social media provided me more opportunity to share relevant material with those asthma patients who have low knowledge as they do more emergency visits and increase work overload during global pandemic.*

Discussion

Principal Findings

The findings show that the increased use of SM has helped to enhance both networking and connections among health professionals outside and within public hospitals. Due to the advent and use of SM sites, management in public hospitals can more quickly share the medical histories of critical patients and can ask for advice from experts in other hospitals. SM promotes knowledge and information exchange among health professionals. As a result, they have become more knowledgeable and competent. There is evidence that various global environmental changes and the increasing consumption of petrol have increased the level of air pollution in Saudi Arabia. Therefore, the number of chronic diseases, especially the number of asthma patients, has increased in recent years. Furthermore, the number of patients with type 2 diabetes has increased because of the high use of red meat and junk foods. These chronic patients have lower levels of education and awareness about chronic diseases, including how to avoid them. Health professionals and the Ministry of Health in Saudi Arabia have started to use YouTube and Twitter platforms to educate patients with chronic diseases. This can reduce work overload, emergency visits, and resource utilization. YouTube videos educate patients about medication, healthy diets, the use of inhalers, exercise plans, and methods to inject insulin to control diabetes.

There is a need to further enhance the productive use of SM as it enhances collaboration and socialization among health professionals and patients. For example, when a health professional is linked to patients with type 2 diabetes through SM, a shared video on YouTube can provide fruitful advice about exercise, the use of medication, the use of inhalers, healthy diets, and other ways to control diabetes. For example, if nurses and doctors are engaged through SM, they can manage emergency situations better, especially during public health crises when all patients cannot receive the same level of care and attention from doctors. Therefore, SM use enhances the level of information sharing. It facilitates learning, knowledge exchange, and patient education about healthy lifestyles. Using SM, patient engagement can enhance the experiences and

XSL•FO
RenderX

knowledge of health professionals, which is useful for improving individual and organizational performance. Furthermore, SM helps to educate and activate patients and can therefore reduce the workload of health organizations as patients are able to control critical situations. Within the organizations in question, the increasing use of SM among nurses and doctors is beneficial for multitasking, as health professionals can ask, guide, or instruct others on routine and emergency tasks. Therefore, it can be argued that the use of SM enhances information sharing, knowledge, skills, and time efficiency for health professionals who are now more productive as a result.

The analysis revealed that Saudi hospitals operate under the control of the Saudi Ministry of Health. Therefore, this ministry can create policies, rules, and regulations that can influence communication, decision making, and the level of productivity of health professionals. It is evident that the use of SM drives collaboration and coordination among health professionals and patients. However, unfortunately, there is no Saudi Health Ministry Policy to positively use these platforms in the workplace. Therefore, when health professionals use these platforms, there is no training or technical, motivational, or organizational support for them. This is counterintuitive, given that SM can save professionals' time, energy, and resources. The findings reveal that the use of SM enhances peer-to-peer support, as newly appointed nurses and young doctors can consult with head nurses and senior doctors through video calls. Their shared knowledge can help them, especially when patients are in emergency situations. However, analysis also suggests that the use of SM has created a culture of quick responses and actions among junior professionals, which ultimately increases the stress and pressure they face to fit more work into limited timeframes. Both nurses and doctors use instant messaging and video calls to increase interpersonal and professional collaboration. As a result, they are able to enhance their level of knowledge, speed of work, and overall organizational productivity. Therefore, it is recommended that the Saudi Ministry of Health and government should specify regulations and policies that can increase the use of SM platforms as evidence-based practices that help health professionals communicate and collaborate to manage critical patients.

A key finding of this study is that the use of SM has promoted the exchange of recorded data and technological collaboration among health professionals. This has improved patient-centric communication, health care services, and the overall level of productivity. For example, diabetes experts can now conduct conference calls with their diabetes patients and advise them on how to clean wounds and stop bleeding in emergency situations. YouTube videos of health experts meant that these patients could self-manage in emergency situations, especially when visiting a hospital is risky. These SM support technological collaboration among different health professionals when it comes to diabetes and asthma health experts (endocrinologists and pulmonologists) who can now arrange video calls to share best practices in terms of medication, diets, and health care plans. This is especially useful for patients with multiple diseases. There is evidence that nurses use SM to share videos to educate first-time parents about infant childcare. Indeed, the data sharing facility provided by SM saves nurses' time so that they can prioritize clinical care and balance it with educating parents and patients.

The use of SM enhances employee engagement by building socialization and trust. It supports shared communication and better patient communication, especially during a global pandemic when health resources are limited, and it is less common for doctors to see patients in person. The findings also reveal that both nurses and doctors engage with instant messaging apps to join appointments. They also use these to share information about promotions, personal educational achievements, parties, and mutual self-help. Therefore, socialization through SM increases social interactions, knowledge exchange, and trust building, and such media can help with emergency situations. Such shared communication is facilitated through SM as doctors and nurses can discuss health care plans and behavioral issues in relation to patients during global pandemics. During global pandemics, both nurses and doctors are unable to provide high standards of health care services because of limited human capital and health resources. However, SM platforms provide them with the opportunity to disseminate educational materials about chronic diseases so that emergency trips to hospitals can be reduced, whereas awareness and self-care can be increased.

Contribution

Previous researchers have applied various models and theories to understand the adoption of technology or SM in the workplace. Examples include the likelihood model [28,29], the technology acceptance model [30-32], the unified theory of acceptance and use of technology 2 [33], the diffusion of innovation theory [34], the social cognitive theory [15], the theory of planned behavior model [35], social exchange theory [3], and social influence theory [25]. However, none of these models have been tested when people are practicing social distancing and are restricted to their homes because of a pandemic. This is the first study to explore how SM practices are developed at the institutional level with the purpose of enhancing technological collaboration among health professionals and patient-centric approaches during the COVID-19 pandemic.

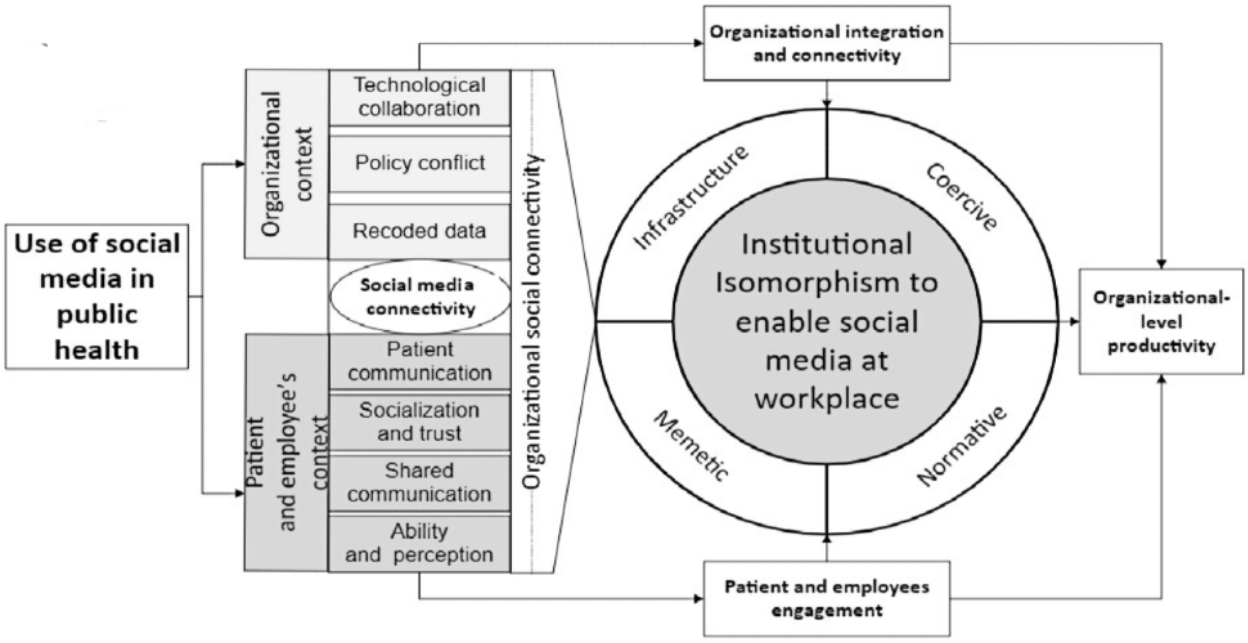
In terms of the first theoretical contribution, this study applied institutional and AMO theories to understand how health professionals experience instant messaging apps. The results could be used to improve professional socialization, trust building, data recording, knowledge exchange, collaboration, health service delivery, and organizational productivity. For example, institutional theory offers knowledge about how to use accessible technology to deal with uncertainty, whereas the ability and motivation of health professionals to use SM provides them with the opportunity to cope with uncertainty during a global pandemic. It empowers them to educate patients and, subsequently, lowers emergency visits.

The second theoretical contribution is the use of institutional theory through the lens of health professionals' motivations and opportunities that are explored in the specific context of Saudi public sector hospitals. For example, institutional theory offers an understanding of professional networking and collaboration, whereas the AMO theory provides an understanding of how

XSL•FO
RenderX

both diabetes and asthma health experts (endocrinologists and pulmonologists) can arrange video calls to share best practices in terms of medication, diet, and health care plans for patients who have multiple diseases and who are vulnerable during a global pandemic. The third and most significant theoretical contribution is that the SMOP model was developed in this study. This model suggests that, even in the absence of institutional policies and in the context of uncertainty, the use of SM platforms at the individual level provides the motivation to develop health professionals and patient interactions with the purpose of creating health care plans during a global pandemic (Figure 3). This model provides an understanding of how to use SM to increase the accessibility of health care professionals at the institutional level. This way, shared communication can be used to manage patients who are at high risk (eg, older and children who live with psychological and behavioral issues).

Figure 3. Social Media Organizational Productivity model.



Practical Implications

This study has provided useful insights regarding how health professionals can make the best use of SM platforms to deal with uncertainty during global pandemics and can reduce the number of emergency visits patients make to hospitals. The second practical contribution is that the use of SM has created a means to record data on how, for example, new parents should care for newborns and how patients with asthma and patients with type 2 diabetes can manage emergency situations. SM can also decrease the workload of professionals during a global pandemic. These data and the technological collaboration between health professionals and patients can reduce the necessity for hospital admissions, and it is useful to ensure the safety of families, communities, and working environments during global pandemics. The third practical contribution is how the use of SM creates an opportunity, regardless of circumstances, to enhance the productive use of SM during a global pandemic. Therefore, it is recommended that the Saudi Ministry of Health and Saudi government should create institutional policies that support the use of SM platforms in public and private hospital settings.

Limitations and Future Directions

Although this study provides unique theoretical and practical contributions, it is not free from limitations. For example, previous studies have highlighted that although qualitative methods can provide rich insights to develop theoretical frameworks, they cannot test the validity of theoretical frameworks [24,25]. Therefore, future studies may test the validity of the SMOP model by collecting data from health professionals in private and public hospitals using statistical methods. Future studies might also test the effectiveness of the SMOP model in different environments and contexts where there are institutional policies and skilled leadership teams that can create more motivation and opportunities to use SM platforms to increase organizational productivity.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic information of doctors.
[DOCX File , 14 KB - jmir_v23i5e23792_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Demographic information of nurses.
[DOCX File , 15 KB - jmir_v23i5e23792_app2.docx]

References

1. Galbin A. An introduction to social constructionism. Social Research Reports. 2014. URL: https://www.researchreports.ro/an-introduction-to-social-constructionism [accessed 2021-04-09]
2. Saudi Arabia is the most Twitter-crazy country in the world: Business Insider. Ministry of Communication and information technology. URL: https://www.mcit.gov.sa/en/media-center/news/91426 [accessed 2020-10-22]
3. Chen X, Wei S. The impact of social media use for communication and social exchange relationship on employee performance. J Knowl Manag 2020 Jun 19;24(6):1289-1314. [doi: 10.1108/jkm-04-2019-0167]
4. Cheng T, Zhang P, Wen Y, Wang L. Social media use and employee innovative performance: work engagement as a mediator. Soc Behav Pers 2020 Jan 07;48(1):1-9. [doi: 10.2224/sbp.8698]
5. Benitez J, Ruiz L, Castillo A, Llorens J. How corporate social responsibility activities influence employer reputation: the role of social media capability. Decis Support Syst 2020 Feb;129:113223. [doi: 10.1016/j.dss.2019.113223]
6. Martinez-Millana A, Fernandez-Llatas C, Bilbao IB, Salcedo MT, Salcedo VT. Evaluating the social media performance of hospitals in Spain: a longitudinal and comparative study. J Med Internet Res 2017 May 23;19(5):e181 [FREE Full text] [doi: 10.2196/jmir.6763] [Medline: 28536091]
7. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 02;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
8. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
9. 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 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
10. Keller B, Labrique A, Jain KM, Pekosz A, Levine O. Mind the gap: social media engagement by public health researchers. J Med Internet Res 2014 Jan 14;16(1):e8 [FREE Full text] [doi: 10.2196/jmir.2982] [Medline: 24425670]
11. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
12. Why spy on Twitter? For Saudi Arabia, it's the Town Square. The New York Times. 2019. URL: https://www.nytimes.com/2019/11/07/world/middleeast/saudi-arabia-twitter-arrests.html [accessed 2019-11-07]
13. Saudi lifting ban on Skype, WhatsApp calls, but will monitor them. Reuters. 2017. URL: https://www.reuters.com/article/us-saudi-telecoms-ban-idUSKCN1BW12L [accessed 2017-09-21]
14. Badea M. Social media and organizational communication. Procedia Soc Behav Sci 2014 Sep;149:70-75. [doi: 10.1016/j.sbspro.2014.08.192]
15. Cao X, Guo X, Vogel D, Zhang X. Exploring the influence of social media on employee work performance. Internet Res 2016. [doi: 10.1108/IntR-11-2014-0299]
16. Huang Y, Singh PV, Ghose A. A structural model of employee behavioral dynamics in enterprise social media. Manage Sci 2015 Dec;61(12):2825-2844. [doi: 10.1287/mnsc.2014.2125]
17. Dong JQ, Wu W. Business value of social media technologies: evidence from online user innovation communities. J Strateg Inf Syst 2015 Jun;24(2):113-127. [doi: 10.1016/j.jsis.2015.04.003]
18. Rueda L, Benitez J, Braojos J. From traditional education technologies to student satisfaction in management education: a theory of the role of social media applications. Inf Manag 2017 Dec;54(8):1059-1071. [doi: 10.1016/j.im.2017.06.002]
19. Zhou L, Zhang D, Yang CC, Wang Y. Harnessing social media for health information management. Electron Commer Res Appl 2018 Jan;27:139-151 [FREE Full text] [doi: 10.1016/j.elerap.2017.12.003] [Medline: 30147636]
20. 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]
21. Allen BJ. Social constructionism. In: Engaging Organizational Communication Theory & Research: Multiple Perspectives. Thousand Oaks, California, United States: SAGE Publications; 2005:35.
22. Taylor B, Francis K. Qualitative Research in the Health Sciences : Methodologies, Methods and Processes. London: Routledge; 2013:1-312.
23. Lock A, Strong T. Social constructionism: sources and stirrings in theory and practice. Cambridge University Press 2010. [doi: 10.1017/CBO9780511815454]
24. Aslam U, Muqadas F, Imran MK, Saboor A. Emerging organizational parameters and their roles in implementation of organizational change. J Organ Chang Manag 2018 Aug 13;31(5):1084-1104. [doi: 10.1108/jocm-08-2017-0300]

XSL•FO
RenderX

25. Naeem M. Developing the antecedents of social influence for internet banking adoption through social networking platforms: evidence from conventional and Islamic banks. Asia Pacific J Mark Logist 2020 Apr 23;33(1):185-204. [doi: 10.1108/apjml-07-2019-0467]
26. Thomas A, Menon A, Boruff J, Rodriguez AM, Ahmed S. Applications of social constructivist learning theories in knowledge translation for healthcare professionals: a scoping review. Implement Sci 2014 May 06;9(1):54 [FREE Full text] [doi: 10.1186/1748-5908-9-54] [Medline: 24885925]
27. Thornton SC, Henneberg SC, Naudé P. Understanding types of organizational networking behaviors in the UK manufacturing sector. Ind Mark Manag 2013 Oct;42(7):1154-1166. [doi: 10.1016/j.indmarman.2013.06.005]
28. Li CY. Persuasive messages on information system acceptance: a theoretical extension of elaboration likelihood model and social influence theory. Comput Hum Behav 2013 Jan;29(1):264-275. [doi: 10.1016/j.chb.2012.09.003]
29. Ngai EW, Tao SS, Moon KK. Social media research: theories, constructs, and conceptual frameworks. J Inf Manag 2015 Feb;35(1):33-44. [doi: 10.1016/j.ijinfomgt.2014.09.004]
30. Al-Maatouk Q, Othman MS, Aldraiweesh A, Alturki U, Al-Rahmi WM, Aljeraiwi AA. Task-technology fit and technology acceptance model application to structure and evaluate the adoption of social media in academia. IEEE Access 2020;8:78427-78440. [doi: 10.1109/access.2020.2990420]
31. Park N, Rhoads M, Hou J, Lee KM. Understanding the acceptance of teleconferencing systems among employees: an extension of the technology acceptance model. Comput Hum Behav 2014 Oct;39:118-127. [doi: 10.1016/j.chb.2014.05.048]
32. Wamba SF, Bhattacharya M, Trinchera L, Ngai EW. Role of intrinsic and extrinsic factors in user social media acceptance within workspace: assessing unobserved heterogeneity. Int J Inf Manag 2017 Apr;37(2):1-13. [doi: 10.1016/j.ijinfomgt.2016.11.004]
33. Ouirdi ME, Ouirdi AE, Segers J, Pais I. Technology adoption in employee recruitment: the case of social media in Central and Eastern Europe. Comput Hum Behav 2016 Apr;57:240-249. [doi: 10.1016/j.chb.2015.12.043]
34. Khan NA, Khan AN. What followers are saying about transformational leaders fostering employee innovation via organisational learning, knowledge sharing and social media use in public organisations? Gov Inf Q 2019 Oct;36(4):101391. [doi: 10.1016/j.giq.2019.07.003]
35. van Zoonen W, Verhoeven JW, Elving WJ. Understanding work-related social media use: an extension of theory of planned behavior. Int J Manag Econo Soc Sci 2014;3(4):164 [FREE Full text]

Abbreviations

AMO: ability, motivation, and opportunity theory
ICU: intensive care unit
SM: social media
SMOP: Social Media Organizational Productivity

Edited by G Eysenbach; submitted 23.08.20; peer-reviewed by U Aslam, F Muqadas; comments to author 11.09.20; revised version received 25.02.21; accepted 03.03.21; published 05.05.21.

Please cite as:
Dailah Sr HG, Naeem M
A Social Media Organizational Productivity Model: Insights From Public Health Professionals
J Med Internet Res 2021;23(5):e23792
URL: https://www.jmir.org/2021/5/e23792
doi:10.2196/23792
PMID:33949965

©Hamad Ghaleb Dailah Sr, Muhammad Naeem. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2021. 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

Sleep Disturbances in Frontline Health Care Workers During the COVID-19 Pandemic: Social Media Survey Study

Nancy H Stewart[1*], DO, MS; Anya Koza[2], BS; Serena Dhaon[3], BS; Christiana Shoushtari[4], MD, MPH; Maylyn Martinez[5*], MD, MPH; Vineet M Arora[5], MD, MAPP

[1]Department of Medicine, University of Kansas Medical Center, Kansas City, KS, United States

[2]Kansas City University, Kansas City, MO, United States

[3]University of Illinois, Chicago, IL, United States

[4]Department of Internal Medicine, Advocate Lutheran General Hospital, Park Ridge, IL, United States

[5]Department of Medicine, University of Chicago Medical Center, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Nancy H Stewart, DO, MS
Department of Medicine
University of Kansas Medical Center
4000 Cambridge St.
Mailstop 3007
Kansas City, KS, 66106
United States
Phone: 1 913 588 6045
Fax: 1 913 588 4098
Email: nstewart5@kumc.edu

Abstract

Background: During the COVID-19 pandemic, health care workers are sharing their challenges, including sleep disturbances, on social media; however, no study has evaluated sleep in predominantly US frontline health care workers during the COVID-19 pandemic.

Objective: The aim of this study was to assess sleep among a sample of predominantly US frontline health care workers during the COVID-19 pandemic using validated measures through a survey distributed on social media.

Methods: A self-selection survey was distributed on Facebook, Twitter, and Instagram for 16 days (August 31 to September 15, 2020), targeting health care workers who were clinically active during the COVID-19 pandemic. Study participants completed the Pittsburgh Sleep Quality Index (PSQI) and Insomnia Severity Index (ISI), and they reported their demographic and career information. Poor sleep quality was defined as a PSQI score ≥5. Moderate-to-severe insomnia was defined as an ISI score >14. The Mini-Z Burnout Survey was used to measure burnout. Multivariate logistic regression tested associations between demographics, career characteristics, and sleep outcomes.

Results: A total of 963 surveys were completed. Participants were predominantly White (894/963, 92.8%), female (707/963, 73.4%), aged 30-49 years (692/963, 71.9%), and physicians (620/963, 64.4%). Mean sleep duration was 6.1 hours (SD 1.2). Nearly 96% (920/963, 95.5%) of participants reported poor sleep (PSQI). One-third (288/963, 30%) reported moderate or severe insomnia. Many participants (554/910, 60.9%) experienced sleep disruptions due to device use or had nightmares at least once per week (420/929, 45.2%). Over 50% (525/932, 56.3%) reported burnout. In multivariable logistic regressions, nonphysician (odds ratio [OR] 2.4, 95% CI 1.7-3.4), caring for patients with COVID-19 (OR 1.8, 95% CI 1.2-2.8), Hispanic ethnicity (OR 2.2, 95% CI 1.4-3.5), female sex (OR 1.6, 95% CI 1.1-2.4), and having a sleep disorder (OR 4.3, 95% CI 2.7-6.9) were associated with increased odds of insomnia. In open-ended comments (n=310), poor sleep was mapped to four categories: children and family, work demands, personal health, and pandemic-related sleep disturbances.

Conclusions: During the COVID-19 pandemic, nearly all the frontline health care workers surveyed on social media reported poor sleep, over one-third reported insomnia, and over half reported burnout. Many also reported sleep disruptions due to device use and nightmares. Sleep interventions for frontline health care workers are urgently needed.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e27331) doi:10.2196/27331

KEYWORDS

social media; sleep disorders; frontline health care workers; burnout; insomnia; sleep; health care worker; stress; survey; demographic; outcome; COVID-19

Introduction

Since March 2020, the pandemic of SARS-CoV-2, the virus that causes COVID-19, has impacted people's lives on a global scale, especially the lives of health care workers on the front lines. During the pandemic, health care workers have used social media to share their experiences serving on the front lines, such as lack of personal protective equipment, mental anguish, burnout [1], and the toll of caring for patients with COVID-19. Repeatedly, health care workers have reported concerns that include sleep loss and other sleep disturbances [2]. Although these sleep issues certainly can be attributed to long working hours, they may also occur for other reasons. Social media posts suggest that health care workers are suffering from nightmares, insomnia, and using mobile devices at night to review social media ("doomscrolling") [3,4]. Given the increased professional and personal responsibilities of health care workers during the pandemic, it is possible that sleep loss or disturbances will be compounded by additional worries during the pandemic. Moreover, mounting evidence is demonstrating that inadequate physician sleep is associated with burnout and medical errors [5,6].

Several studies in various countries have described the sleep of health care workers during the COVID-19 pandemic. Most notable are two meta-analyses on sleep in health care workers during the initial months of the COVID-19 pandemic, with studies mostly focusing on the experience of health care workers in China. Pappa et al [7] reviewed 5 cross-sectional studies conducted among Chinese health care workers prior to April 17, 2020, and reported a prevalence of insomnia (38.9%) among them. Salari et al [8] reviewed 7 cross-sectional studies from Asia and the Middle East conducted among nurses and physicians prior to June 24, 2020, and they reported a prevalence of sleep disturbances of 34.8% and 41.6% in nurses and physicians, respectively. Additionally, several other studies reported on sleep loss in health care workers from countries such as Oman [9], Bahrain [10], and Spain [11].

Prior to the COVID-19 pandemic, studies reported that nearly 1 in 2 physicians admitted experiencing symptoms of burnout [12]. In 2004, researchers described the profound impact of perceived sleep loss on trainee well-being [13]. That same year, Lockley et al [14] demonstrated that reduction of intern work hours improved sleep and decreased attention failures. In 2015, the American Thoracic Society published an official statement affirming that good quality and quantity of sleep is imperative for healthy physicians [15].

To date, few studies of the sleep experience among frontline health care workers during the COVID-19 pandemic in the United States have been conducted. These studies did not utilize validated measures of sleep and were limited to a single institution [3,16]. In the United States, to our knowledge, no study has yet evaluated sleep using standardized measures in clinically active frontline health care workers during the COVID-19 pandemic. Despite anecdotal observations of sleep loss described via social media platforms, no study has yet explicitly explored the sleep of frontline health care workers during the COVID-19 pandemic utilizing social media, an increasingly important and influential mode of communication during the pandemic. The aim of this study was to describe the sleep disturbances of a sample of predominantly US frontline health care workers during the COVID-19 pandemic with validated instruments through a survey distributed on social media. A secondary aim of this study was to examine the association between the demographic and career characteristics of health care workers and their reported sleep disturbances and burnout.

Methods

From August 31 to September 15, 2020, we recruited frontline health care workers from across the world to complete the web-based R.E.S.T. (Recommending Essential Sleep Time) survey. The restricted, self-selection survey (respondents chose to participate by clicking on the survey link) was distributed through Twitter, Facebook, and Instagram social media accounts created specifically for the study. A river sampling method was used to recruit participants anonymously [17]. This web-based sampling technique redirected interested participants to the survey link via digital promotion (eg, banners, advertisements, flyers, offers) [17]. A unique bit.ly link was created for the social media accounts to enable tracking of views and clicks. Participation in the study was voluntary, and participants could opt out at any time while responding to the survey questions. Participants included physicians, physician assistants, nurse practitioners, nurses, pharmacists, physical/ occupational/ respiratory therapists, health care trainees, and allied health care workers (eg, medical technologists, therapists, certified nursing assistants). The University of Chicago Institutional Review Board reviewed and approved the protocol.

Survey Design

Study data were collected and managed using Research Electronic Data Capture (REDCap), hosted at the University of Chicago. REDCap is a secure, web-based software platform that supports data capture for research studies [18,19]. Qualifying participants for our study were at least 18 years of age and were frontline health care workers who were clinically active during the COVID-19 pandemic. The survey automatically terminated for respondents who did not meet these inclusion criteria. We asked the participants to frame all their responses for 1 month in the pandemic in which they experienced the greatest clinical intensity and risk of COVID-19 transmission. This time frame was selected as the most representative of the sleep of respondents while they were

XSL•FO
RenderX

clinically active on the front lines caring for patients with COVID-19.

The survey contained the following categories: Sleep, Well-being, Career, and Baseline Demographics. The Sleep section included the Pittsburgh Sleep Quality Index (PSQI) [20] and the Insomnia Severity Index (ISI) [21]. The PSQI is a 19-item validated questionnaire that measures sleep quality over 1 month, generating 1 global score [20]. The ISI is a 7-item validated questionnaire that assesses the nature, severity, and impact of insomnia in adults [21]. The Sleep section also included questions about sleep disturbances due to device use. The Well-being section included the Mini-Z Burnout Survey, an 8-item validated questionnaire that investigates health care burnout and job satisfaction [22]. The Mini-Z also includes a 1-item burnout measure that is validated against the longer Maslach Burnout Inventory [23]. Other questions in this section related to pandemic-related worries, changing responsibilities, living situation, and health status. The Career section collected data on each participant's job role, level of training, specialty, and career setting prior to the pandemic and whether they cared for patients with COVID-19 directly. Baseline demographics were collected on marital status, age, gender, race, ethnicity, type of practice, location of practice, and type of community served (eg, urban, suburban, rural). Two open-ended questions were also included in the survey: "What was the cause of your sleep disruption(s) during your reported month?" and "Is there anything else you would like to share?"

Upon completion of the survey, participants were invited to submit their email address to enter a daily raffle for a US $25 Amazon gift card (14 gift cards in total). Before deployment, the survey was piloted with a few health care workers who were not affiliated with the study. Changes made included deleting duplicate questions and removing one sleep scale to abbreviate the time to complete the survey to 10 minutes or less.

Social Media Distribution Strategy

To disseminate the survey link, we created a Facebook page and a Twitter and Instagram account on behalf of the R.E.S.T. study. This enabled the use of engagement tracking built into these social media platforms to calculate the response rate. To market our survey through Instagram, we used paid advertisements. On Facebook, we posted on health care–specific social media groups (eg, Women Physician Wellness, StyleMD, and the Physician Collective). On Twitter, we tagged colleagues and organizations and employed health care–specific hashtags (eg, #MedTwitter, #NurseTwitter). Although the authors re-shared the R.E.S.T. survey posts on their personal Twitter accounts to increase visibility, the link remained connected to the original posts of the R.E.S.T. accounts to preserve the ability to track engagement. To promote the survey, an infographic was posted each day on the study's social media accounts. Survey links were also placed in the biography of each social media account. To draw attention to the survey post and the account, we linked each post to standard information from national organizations (eg, the American Academy of Sleep Medicine and the Sleep Research Society) and to articles about sleep and the pandemic through the study's social media accounts.

One of our study authors (SRD) used a customized *bit.ly* link [24] to track user engagement daily on Facebook, Twitter, Instagram, and email. The metrics analytics from the social media platforms enabled us to view specific metrics on reach (the total number of individuals who viewed the content) and link clicks (the number of times individuals chose to open the survey). At midnight on each day, we calculated the reach and link clicks of the study on each social media website. This enabled us to compare the performance of each platform and track our progress. We also calculated the number of survey responses per day while the survey was active.

Statistical Analysis

Descriptive statistics were used to quantify the participants' demographics and profession. Primary outcomes were poor sleep quality and moderate to severe insomnia. Poor sleep quality was defined as a PSQI score ≥5 [20]. Moderate-to-severe insomnia was defined as an ISI score of >14.21. The Mini-Z Burnout Survey was used to measure burnout [22]. Multivariate logistic regression was used to test for independent associations between age, gender, race (Black vs non-Black), ethnicity (Hispanic vs non-Hispanic), and profession (physician vs. nonphysician), with odds of each of the outcomes, after controlling for those with a preexisting sleep disorder. All statistical analyses were conducted using Stata Statistical Software, Release 16 (StataCorp LLC), with $P<.05$ used to indicate statistical significance.

Coding Open-ended Comments

There were two open-ended questions in the survey. For participants who responded to one or both questions, open coding of comments was used [25]. Sleep disturbance was mapped to four main themes: (1) children and family; (2) work demands affecting sleep; (3) personal health conditions; and (4) pandemic-related sleep disturbances. Subthemes were created for each of these themes.

Results

Our social media posts were seen by 87,061 unique individuals and resulted in 976 clicks on our survey link. Our final sample contained 963 submitted surveys. The most common ways people discovered the survey were through Twitter (372/944, 39.4%), in a private Facebook group (199/944, 21.1%), and through a colleague (103/944, 10.9%).

Participants were mostly female (707/963, 73.4%), White (894/963, 92.8%), aged 30-49 years (692/963, 71.9%), and physicians (620/963, 64.4%) (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the sample of survey respondents (N=963).

Characteristic	Value, n (%)
Age group (years)	
18-29	149 (15.5)
30-49	692 (71.9)
50+	86 (8.9)
Did not answer	36 (3.7)
Gender	
Male	220 (22.9)
Female	707 (73.4)
Did not answer	36 (3.7)
Race[a]	
White	894 (92.8)
Black	69 (7.2)
Ethnicity[b]	
Non-Hispanic	759 (78.8)
Hispanic	124 (12.9)
Did not answer	80 (8.3)
Significant other	
Yes	700 (72.7)
No	220 (22.9)
Did not answer	43 (4.5)
Career role	
Physician	620 (64.4)
Nonphysician	343 (35.6)
Directly cared for patients with COVID-19	744 (80)
Country or continent of residence	
United States	882 (91.6)
Canada	18 (1.9)
Central America	10 (1.0)
South America	9 (0.9)
Europe	7 (0.7)
Asia/Oceania	4 (0.4)
Australia/New Zealand	4 (0.4)
Africa	3 (0.3)
Did not answer	26 (2.7)

[a]Reference group for race: non-Black.

[b]Reference group for ethnicity: non-Hispanic.

Poor sleep quality, defined as a PSQI score ≥5, was identified in 95.5% (920/963) of all health care workers (Table 2). Rates of moderate and severe insomnia were 249/963 (28.5%) and 39/963 (4.5%), respectively (Table 2). Average sleep duration was 6.1 hours per day (SD 1.2) (Table 2).

Many respondents (554/910, 60.9%) reported sleep disruptions due to device use or due to nightmares at least once per week (420/929, 45%). Those with a sleep disorder were more likely to report burnout (52.1% vs 63.2%, χ^2_4=116.4; P<.001). In unadjusted analyses, being Black (31.0% vs 59.4%), having Hispanic ethnicity (27.9% vs 59.6%), being single (44.6% vs 27.1%), and being a non-physician (50.7% vs 23.2%) were significantly associated with risk of moderate-to-severe insomnia. Non-Hispanic ethnicity was the reference group. In

XSL•FO
RenderX

the multivariable logistic regressions, not being a physician (odds ratio [OR] 2.4, 95% CI: 1.7-3.4), caring for patients with COVID-19 (OR 1.8, 95% CI 1.2-2.8), Hispanic ethnicity (OR 2.2, 95% CI 1.4-3.5), and being female (OR 1.6, 95% CI 1.1-2.4) were associated with increased odds of insomnia (Table 3). Having a sleeping disorder also increased the odds of insomnia compared to those without one (OR 4.3, 95% CI 2.7-6.9) (Table 3). With respect to confounders, we did not focus on low socioeconomic status; however, we did evaluate geography, and we noted no major impact on geography.

Table 2. Prevalence of poor sleep and insomnia in health care workers working during the COVID-19 pandemic (N=963).

Variable	Value
Poor sleep quality[a], n (%)	920 (95.5)
Insomnia[b], n (%)	
Absent	220 (25.1)
Subthreshold	367 (41.9)
Moderate	249 (28.5)
Severe	39 (4.5)
Sleep duration (hours)[c], average (SD)	6.1 (1.2)

[a]As measured by the Pittsburgh Sleep Quality Index (PSQI); poor sleep quality was defined as a PSQI score ≥5 [20].

[b]As measured by the Insomnia Severity Index (ISI); insomnia was qualified by ISI score, interpreted as absent (0-7), subthreshold (8-14), moderate (15-21), or severe (22-28) [21].

[c]Average hours of sleep duration was defined by the PSQI [20].

Table 3. Risk factors for insomnia among frontline health care workers during the COVID-19 pandemic (N=963).

Characteristic	Insomnia[a], odds ratio (95% CI)	*P* value
Age group (years)		
18-29	Reference	N/A[b]
30-49	1.1 (0.7-1.8)	.63
≥50	1.3 (0.7-2.4)	.49
Gender		
Male	Reference	N/A
Female	1.6 (1.1-2.4)	.02
Black race	1.5 (0.8-2.7)	.24
Hispanic ethnicity	2.2 (1.4-3.5)	<.001
No significant other	1.3 (0.9-1.9)	.18
Career role		
Physician	Reference	N/A
Nonphysician	2.4 (1.7-3.4)	<.001
Prior sleep disorder	4.3 (2.7-6.9)	<.001
Cared for patients with COVID-19	1.8 (1.2-2.8)	.008

[a]As measured by the Insomnia Sleep Index (ISI); insomnia was defined by ISI scores qualifying as moderate to severe insomnia (>14) [21].

[b]N/A: not applicable.

A total of 310 open-ended comments regarding factors affecting sleep were categorized into the following four themes: (1) children and family; (2) work demands affecting sleep; (3) personal health conditions; and (4) pandemic-related sleep disturbances. The most frequently reported non–pandemic-related theme was children and family (n=59) (“COVID plus home stress plus stress over my kids, my job, my marriage”) (Table 4). Other non–pandemic-related sleep disturbances included work demands affecting sleep (n=48) (“The volume of calls and messages from my patient and caregiver population is through the roof and I’m sleeping 4-5 hours per night”) and personal health (n=41) (“Insomnia predating COVID, but worsened with COVID”) (Table 4). Pandemic-related sleep disturbances (n=48) (“I never had sleep issues prior to the COVID-19 pandemic; suddenly I had issues with sleep initiation”) were also noted (Table 4).

XSL•FO
RenderX

Table 4. Themes, subthemes, and example quotes from the open-ended comments (n=310).

Theme	Subthemes	n (%)	Example quote
Sleep disruptions not due to the pandemic			
Children and family (n=59)	Child care (1-18 years)	27 (46)	"6 year old who wakes 2-3 times per night (some secondary to COVID related anxiety for him), 2.5 year old who wakes about 2-3 nights per week"
	Infant care	18 (31)	"Both COVID and my usual sleep disturbances plus my infant daughter"
	Concerns about family affecting sleep	8 (17)	"COVID plus home stress plus stress over my kids, my job, my marriage"
	Pregnancy affecting sleep	6 (10)	"I was pregnant with twins and carried to term. COVID increased my anxiety and work-related stress. I had to lay off employees and reduce my salary, right before maternity leave."
Work demands affecting sleep (n=48)	Shift work	18 (38)	"Mostly schedule fluctuations each week - late evening shifts followed by early morning shifts"
	Work-life interference	18 (38)	"Extra work at home [charting] due to increased patient volumes"
	Residency training	6 (13)	"Switching schedules with residency (days/nights and 28-hour calls), overall sleep debt, and fatigue"
	Volume of calls/pages	6 (13)	"The volume of calls and messages from my patient and caregiver population is through the roof…sleeping 4-5 hours [per night]."
Personal health conditions (n=41)	Poor sleep prior to the pandemic	18 (44)	"I have never been a great sleeper."
	Chronic medical issue	12 (29)	"Possibly COVID-19 related.... I had an onset of Obsessive-Compulsive Disorder this summer and have been adjusting to antipsychotic medication, intervention, and prescription sleep aids."
	Formal sleep disorder	8 (20)	"Insomnia pre-dating COVID, but [it] worsened with COVID"
	Female hormones	3 (7)	"In vitro fertilization side effects"
Sleep disruptions due to the pandemic			
Pandemic-related sleep disturbances (n=48)	Pandemic affecting sleep	17 (35)	"I never had sleep issues prior to the COVID-19 pandemic. Suddenly I had issues with sleep initiation."
	Nightmares about the pandemic	8 (17)	"The only thing that has changed with my sleep is an increase in nightmares and night terrors primarily regarding work situations."
	Pandemic-related exhaustion	8 (17)	"In the last 1-2 months, I am EXHAUSTED and fall asleep easily, often on the couch right after dinner."
	Medication/substance use to cope with the pandemic	7 (15)	"Alcohol plays a role in my sleep deprivation and 'coping' during COVID."
	Worry about pandemic response	4 (8)	"I worry about how COVID is being managed by the President of the United States and I don't believe he is doing a good job; I worry things will get even worse if he is elected to a second term. This does keep me awake at night."

Discussion

Principal Findings

To our knowledge, this is the first study to evaluate self-reported sleep disturbances among a sample of predominantly US frontline health care workers during the COVID-19 pandemic with validated instruments using a survey distributed on social media. Our study demonstrates that nearly all (96%) of the surveyed health care workers reported poor sleep, over 30% reported moderate to severe insomnia, and over 50% reported burnout. A majority of respondents also reported sleep disruptions due to personal technology device use (eg, cell phone, iPad) or nightmares at least once per week. Additionally, survey participants reported a mean sleep duration of 6.1 hours per day, which is less than the recommended 7 hours of sleep for US adults [26]. Our study demonstrated that health care workers at highest risk of insomnia included those who were not physicians, were not male, were Hispanic, and directly cared for patients with COVID-19. We also demonstrate that sleep disturbances in health care workers during the pandemic are related not only to work demands and personal health but also to children and family as well as to the pandemic itself, including worry about the response to the COVID-19 pandemic. Given the strong association between sleep and burnout among health professionals, it is not surprising that a high number of the surveyed health care workers reported burnout [5].

XSL•FO
RenderX

The results of this study are consistent with prior studies demonstrating a high prevalence of insomnia in nurses and physicians caring for patients with COVID-19 on the front lines [8]. In a systematic review and meta-analysis, Salari et al [8] reported on the increase of insomnia in frontline nurses near 35% and in frontline physicians near 42%, however they did not report on a subgroup analysis of insomnia in women. In the meta-analysis by Pappa et al [7], the prevalence of insomnia in health care workers during the COVID-19 pandemic was near 40%, and the authors did not specifically report the prevalence of insomnia in women frontline health care workers. Our data are also in accord with those in a study by Trockel et al [5], who demonstrated that sleep-related impairment is an occupational hazard associated with increased medical-related errors and burnout. In comparing our sample to the general US physician population, we found higher rates of sleep disturbances and insomnia risk than in the general population of practicing physicians [12,27,28]. Our study findings of sleep disruption due to device use are consistent with findings demonstrating increased nightly screen time in health care workers and reports of doomscrolling [3,4].

Our findings have many clinical implications. Frontline health care workers involved in the direct care of patients with COVID-19 report decreased sleep time, increased insomnia, nightmares, fears for their safety, increased clinical workload, and concerns for their family, which reportedly impact their sleep. Health care workers are at high risk for the development of psychological distress and medical errors [5,29]. The professional environment for health care workers has drastically changed during the pandemic, which has brought challenges related to increased workload, reduced protective equipment and resources, rapidly evolving protocols, relocation of intensive care settings, fear of viral transmission, and social isolation from supportive networks [30,31]. Given this unfolding environment, the findings of sleep disturbances of frontline health care workers are not surprising.

The study findings are particularly concerning given the natural course of insomnia and potential for long-term psychological impact, as seen in health care workers during prior severe acute respiratory syndrome (SARS) outbreaks [32-34]. In a recently reported longitudinal study on insomnia, nearly half (42%) of patients who had insomnia initially (as measured by the ISI) also had insomnia 5 years later, demonstrating a persistent course [35]. Likewise, studies on the mental health of health care workers who were active during the 2003 SARS epidemic suggest the psychological consequences may persist months to years after the epidemic ends [34]. While noting long-term impacts of mental health, it is also important to consider the implication of potential medical errors due to sleep loss by these health care workers [6]. Studies of sleep disorders in health care workers during prior infectious outbreaks are in concert with our study findings, and they forecast the potential longitudinal psychological impact on the sleep health of our frontline health care colleagues.

Limitations

Although social media platforms are widely used, a limitation of this study may be that a sample recruited through social media platforms may not be representative of health care workers in the US and around the world. To combat time period bias, we asked participants to complete the survey while thinking of the month in which they were most clinically active in caring for patients with COVID-19. Although this approach may lead to biased results, we felt it would provide results that were most representative of the respondents' experiences while they were serving on the front lines during the COVID-19 pandemic. Due to the nature of a worldwide pandemic, the incidence of COVID-19 is distributed differently across geographic areas; therefore, the time frame of reflection may differ between respondents. Because the focus of our survey was on sleep, it is possible that health care workers who were more likely to experience sleep disturbances completed the survey. Given the prevalence of sleep disorders in the general US population of 8%-19% [36,37], our sample does not appear to overrepresent people with sleep disorders. Our sample did not contain a sufficient number of nonphysician health care workers to characterize the sleep quality of nonphysician workers (eg, physician assistants, nurse practitioners, nurses, pharmacists, allied health workers) during the COVID-19 pandemic. We also asked the participants to complete the survey questions by reflecting on a time when they experienced the greatest clinical intensity and risk of COVID-19 transmission; this may have led to inconsistencies in our data. All data were self-reported, and we do not have objective measures of sleep duration and quality in this sample. We also do not have longitudinal data on our sample, limiting our ability to draw longitudinal conclusions. Furthermore, when using data obtained from social media, it is difficult to capture the accurate outreach and link clicks; for example, if an individual shares the link but not our original post, we cannot calculate the outreach and metrics. Therefore, the actual metrics may be higher than we were capable of measuring.

Prior to the pandemic, sleep and burnout issues existed among health care workers; however, strategies to improve and implement change are lagging behind. The reporting of these issues by frontline health care workers caring for patients with COVID-19 during the pandemic is not surprising. Studies demonstrating successful interventions to impact sleep and burnout are increasing, and they recommend interventions focused both at the individual and the organizational level. Examples of these recommendations include reducing work hours while maintaining the same salary, allowing schedule flexibility, behavioral wellness, and curricular interventions [38-41].

Conclusion

During the COVID-19 pandemic, almost all the health care workers we surveyed on social media reported poor sleep, with nearly half of respondents reporting moderate to severe insomnia and over half reporting burnout. Immediate interventions to improve the sleep and well-being of health care workers, discourage nocturnal social media use, and discourage doomscrolling are needed to strengthen the ability of these workers to continuously meet the daily demands of the COVID-19 pandemic. Concerns regarding clinician well-being are not new. Moreover, health systems and national organizations are placing greater emphasis on systemic change

XSL•FO
RenderX

[38]. Practical frameworks for creating wellness exist; however, senior-level champions are critical for implementation [42]. Considering that sleep is a modifiable factor of our physical and mental health, prioritizing it may help mitigate the health risks associated with the pandemic and its potential longitudinal impacts on sleep and health.

Conflicts of Interest

None declared.

References

1. Dubey S, Biswas P, Ghosh R, Chatterjee S, Dubey MJ, Chatterjee S, et al. Psychosocial impact of COVID-19. Diabetes Metab Syndr 2020 Sep;14(5):779-788 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.035] [Medline: 32526627]
2. Campbell L. Weed and seven Benadryl: the wild lengths COVID docs are taking to get sleep. Daily Beast. 2021 Jan 03. URL: https://www.thedailybeast.com/weed-and-seven-benadryl-and-other-wild-lengths-covid-docs-are-taking-to-get-sleep [accessed 2021-01-03]
3. Conroy DA, Hadler NL, Cho E, Moreira A, MacKenzie C, Swanson LM, et al. The effects of COVID-19 stay-at-home order on sleep, health, and working patterns: a survey study of US health care workers. J Clin Sleep Med 2021 Feb 01;17(2):185-191. [doi: 10.5664/jcsm.8808] [Medline: 32975194]
4. Paulsen P, Fuller D. Scrolling for data or doom during COVID-19? Can J Public Health 2020 Aug 08;111(4):490-491 [FREE Full text] [doi: 10.17269/s41997-020-00376-5] [Medline: 32642969]
5. Trockel MT, Menon NK, Rowe SG, Stewart MT, Smith R, Lu M, et al. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. JAMA Netw Open 2020 Dec 01;3(12):e2028111 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28111] [Medline: 33284339]
6. Fahrenkopf AM, Sectish TC, Barger LK, Sharek PJ, Lewin D, Chiang VW, et al. Rates of medication errors among depressed and burnt out residents: prospective cohort study. BMJ 2008 Feb 07;336(7642):488-491. [doi: 10.1136/bmj.39469.763218.be]
7. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
8. Salari N, Khazaie H, Hosseinian-Far A, Ghasemi H, Mohammadi M, Shohaimi S, et al. The prevalence of sleep disturbances among physicians and nurses facing the COVID-19 patients: a systematic review and meta-analysis. Global Health 2020 Sep 29;16(1):92 [FREE Full text] [doi: 10.1186/s12992-020-00620-0] [Medline: 32993696]
9. Badahdah AM, Khamis F, Al Mahyijari N. Sleep quality among health care workers during the COVID-19 pandemic. J Clin Sleep Med 2020 Sep 15;16(9):1635-1635. [doi: 10.5664/jcsm.8624] [Medline: 32515347]
10. Jahrami H, BaHammam AS, AlGahtani H, Ebrahim A, Faris M, AlEid K, et al. The examination of sleep quality for frontline healthcare workers during the outbreak of COVID-19. Sleep Breath 2021 Mar 26;25(1):503-511 [FREE Full text] [doi: 10.1007/s11325-020-02135-9] [Medline: 32592021]
11. Herrero San Martin A, Parra Serrano J, Diaz Cambriles T, Arias Arias EM, Muñoz Méndez J, Del Yerro Álvarez MJ, et al. Sleep characteristics in health workers exposed to the COVID-19 pandemic. Sleep Med 2020 Nov;75:388-394 [FREE Full text] [doi: 10.1016/j.sleep.2020.08.013] [Medline: 32950884]
12. Stewart NH, Arora VM. The impact of sleep and circadian disorders on physician burnout. Chest 2019 Nov;156(5):1022-1030 [FREE Full text] [doi: 10.1016/j.chest.2019.07.008] [Medline: 31352036]
13. Papp KK, Stoller EP, Sage P, Aikens JE, Owens J, Avidan A, et al. The effects of sleep loss and fatigue on resident-physicians: a multi-institutional, mixed-method study. Acad Med 2004 May;79(5):394-406. [doi: 10.1097/00001888-200405000-00007] [Medline: 15107278]
14. Lockley SW, Cronin JW, Evans EE, Cade BE, Lee CJ, Landrigan CP, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004 Oct 28;351(18):1829-1837. [doi: 10.1056/nejmoa041404]
15. Mukherjee S, Patel SR, Kales SN, Ayas NT, Strohl KP, Gozal D, American Thoracic Society ad hoc Committee on Healthy Sleep. An official American Thoracic Society statement: the importance of healthy sleep. Recommendations and future priorities. Am J Respir Crit Care Med 2015 Jun 15;191(12):1450-1458 [FREE Full text] [doi: 10.1164/rccm.201504-0767ST] [Medline: 26075423]
16. Shechter A, Diaz F, Moise N, Anstey DE, Ye S, Agarwal S, et al. Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. Gen Hosp Psychiatry 2020 Sep;66:1-8 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.06.007] [Medline: 32590254]
17. Fricker R. Sampling methods for web and e-mail surveys. In: The SAGE Handbook of Online Research Methods. Thousand Oaks, CA: SAGE Publications, Ltd; 2008:195-216.
18. 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]

XSL•FO
RenderX

19. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
20. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Research 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
21. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
22. 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 2;31(9):1004-1010 [FREE Full text] [doi: 10.1007/s11606-016-3720-4] [Medline: 27138425]
23. Rohland BM, Kruse GR, Rohrer JE. Validation of a single-item measure of burnout against the Maslach Burnout Inventory among physicians. Stress Health 2004 Apr 07;20(2):75-79. [doi: 10.1002/smi.1002]
24. R.E.S.T. Survey. URL: https://redcap.uchicago.edu/surveys/?s=4PFAAD8E9C [accessed 2021-05-03]
25. Khandkar SH. Open Coding. University of Calgary - Department of Computer Science. 2009. URL: http://pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/open-coding.pdf [accessed 2021-05-05]
26. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Non-Participating Observers, American Academy of Sleep Medicine Staff, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. J Clin Sleep Med 2015 Jun 15;11(6):591-592 [FREE Full text] [doi: 10.5664/jcsm.4758] [Medline: 25979105]
27. Vela-Bueno A, Moreno-Jiménez B, Rodríguez-Muñoz A, Olavarrieta-Bernardino S, Fernández-Mendoza J, De la Cruz-Troca JJ, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. J Psychosom Res 2008 Apr;64(4):435-442. [doi: 10.1016/j.jpsychores.2007.10.014] [Medline: 18374744]
28. Nishimura K, Nakamura F, Takegami M, Fukuhara S, Nakagawara J, Ogasawara K, et al. Cross-sectional survey of workload and burnout among Japanese physicians working in stroke care. Circ Cardiovasc Qual Outcomes 2014 May;7(3):414-422. [doi: 10.1161/circoutcomes.113.000159]
29. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
30. Anagiotos A, Petrikkos G. Otolaryngology in the COVID-19 pandemic era: the impact on our clinical practice. Eur Arch Otorhinolaryngol 2021 Mar 23;278(3):629-636 [FREE Full text] [doi: 10.1007/s00405-020-06161-x] [Medline: 32577905]
31. Mathiesen T, Arraez M, Asser T, Balak N, Barazi S, Bernucci C, EANS Ethico-legal committee. A snapshot of European neurosurgery December 2019 vs. March 2020: just before and during the Covid-19 pandemic. Acta Neurochir (Wien) 2020 Sep;162(9):2221-2233 [FREE Full text] [doi: 10.1007/s00701-020-04482-8] [Medline: 32642834]
32. Jansson-Fröjmark M, Linton SJ. The course of insomnia over one year: a longitudinal study in the general population in Sweden. Sleep 2008 Jun;31(6):881-886 [FREE Full text] [doi: 10.1093/sleep/31.6.881] [Medline: 18548834]
33. Su TP, Lien TC, Yang CY, Su YL, Wang JH, Tsai SL, et al. Prevalence of psychiatric morbidity and psychological adaptation of the nurses in a structured SARS caring unit during outbreak: a prospective and periodic assessment study in Taiwan. J Psychiatr Res 2007 Jan;41(1-2):119-130 [FREE Full text] [doi: 10.1016/j.jpsychires.2005.12.006] [Medline: 16460760]
34. Luo Y, Chua CR, Xiong Z, Ho RC, Ho CSH. A systematic review of the impact of viral respiratory epidemics on mental health: an implication on the coronavirus disease 2019 pandemic. Front Psychiatry 2020 Nov 23;11:565098 [FREE Full text] [doi: 10.3389/fpsyt.2020.565098] [Medline: 33329106]
35. Morin CM, Jarrin DC, Ivers H, Mérette C, LeBlanc M, Savard J. Incidence, persistence, and remission rates of insomnia over 5 years. JAMA Netw Open 2020 Nov 02;3(11):e2018782 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18782] [Medline: 33156345]
36. Ford ES, Cunningham TJ, Giles WH, Croft JB. Trends in insomnia and excessive daytime sleepiness among U.S. adults from 2002 to 2012. Sleep Med 2015 Mar;16(3):372-378 [FREE Full text] [doi: 10.1016/j.sleep.2014.12.008] [Medline: 25747141]
37. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
38. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov;388(10057):2272-2281. [doi: 10.1016/s0140-6736(16)31279-x]
39. Johnson KM, Simon N, Wicks M, Barr K, O'Connor K, Schaad D. Amount of sleep, daytime sleepiness, hazardous driving, and quality of life of second year medical students. Acad Psychiatry 2017 Oct 18;41(5):669-673. [doi: 10.1007/s40596-017-0668-6] [Medline: 28421480]
40. Schiller H, Lekander M, Rajaleid K, Hellgren C, Åkerstedt T, Barck-Holst P, et al. The impact of reduced worktime on sleep and perceived stress - a group randomized intervention study using diary data. Scand J Work Environ Health 2017 Mar 01;43(2):109-116 [FREE Full text] [doi: 10.5271/sjweh.3610] [Medline: 27942734]

XSL•FO
RenderX

41. Tucker P, Bejerot E, Kecklund G, Aronsson G, Åkerstedt T. The impact of work time control on physicians' sleep and well-being. Appl Ergon 2015 Mar;47:109-116. [doi: 10.1016/j.apergo.2014.09.001] [Medline: 25479980]
42. Adibe B, Perticone K, Hebert C. Creating wellness in a pandemic: a practical framework for health systems responding to Covid-19. NEJM Catalyst. 2021 Mar 13. URL: https://catalyst.nejm.org/doi/pdf/10.1056/CAT.20.0218 [accessed 2021-03-13]

Abbreviations

ISI: Insomnia Severity Index
OR: odds ratio
PSQI: Pittsburgh Sleep Quality Index
R.E.S.T.: Recommending Essential Sleep Time
REDCap: Research Electronic Data Capture

Edited by C Basch; submitted 21.01.21; peer-reviewed by MA Bahrami, D Huang; comments to author 18.02.21; revised version received 14.03.21; accepted 16.04.21; published 19.05.21.

Please cite as:
Stewart NH, Koza A, Dhaon S, Shoushtari C, Martinez M, Arora VM
Sleep Disturbances in Frontline Health Care Workers During the COVID-19 Pandemic: Social Media Survey Study
J Med Internet Res 2021;23(5):e27331
URL: https://www.jmir.org/2021/5/e27331
doi:10.2196/27331
PMID:33875414

©Nancy H Stewart, Anya Koza, Serena Dhaon, Christiana Shoushtari, Maylyn Martinez, Vineet M Arora. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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 an Automatic System for Classifying Chatter About Health Services on Twitter: Case Study for Medicaid

Yuan-Chi Yang[1], PhD; Mohammed Ali Al-Garadi[1], PhD; Whitney Bremer[1], BS; Jane M Zhu[2], MD, MPP, MSHP; David Grande[3], MD, MPA; Abeed Sarker[1,4], PhD

[1]Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States

[2]Division of General Internal Medicine and Geriatrics, Oregon Health & Science University, Portland, OR, United States

[3]Division of General Internal Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, United States

Corresponding Author:
Yuan-Chi Yang, PhD
Department of Biomedical Informatics
School of Medicine
Emory University
101 Woodruff Circle, 4th Floor East
Atlanta, GA, 30322
United States
Phone: 1 404 727 6123
Email: yuan-chi.yang@emory.edu

Abstract

Background: The wide adoption of social media in daily life renders it a rich and effective resource for conducting near real-time assessments of consumers' perceptions of health services. However, its use in these assessments can be challenging because of the vast amount of data and the diversity of content in social media chatter.

Objective: This study aims to develop and evaluate an automatic system involving natural language processing and machine learning to automatically characterize user-posted Twitter data about health services using Medicaid, the single largest source of health coverage in the United States, as an example.

Methods: We collected data from Twitter in two ways: via the public streaming application programming interface using Medicaid-related keywords (Corpus 1) and by using the website's search option for tweets mentioning agency-specific handles (Corpus 2). We manually labeled a sample of tweets in 5 predetermined categories or *other* and artificially increased the number of training posts from specific low-frequency categories. Using the manually labeled data, we trained and evaluated several supervised learning algorithms, including support vector machine, random forest (RF), naïve Bayes, shallow neural network (NN), k-nearest neighbor, bidirectional long short-term memory, and bidirectional encoder representations from transformers (BERT). We then applied the best-performing classifier to the collected tweets for postclassification analyses to assess the utility of our methods.

Results: We manually annotated 11,379 tweets (Corpus 1: 9179; Corpus 2: 2200) and used 7930 (69.7%) for training, 1449 (12.7%) for validation, and 2000 (17.6%) for testing. A classifier based on BERT obtained the highest accuracies (81.7%, Corpus 1; 80.7%, Corpus 2) and F_1 scores on consumer feedback (0.58, Corpus 1; 0.90, Corpus 2), outperforming the second best classifiers in terms of accuracy (74.6%, RF on Corpus 1; 69.4%, RF on Corpus 2) and F_1 score on consumer feedback (0.44, NN on Corpus 1; 0.82, RF on Corpus 2). Postclassification analyses revealed differing intercorpora distributions of tweet categories, with political (400778/628411, 63.78%) and consumer feedback (15073/27337, 55.14%) tweets being the most frequent for Corpus 1 and Corpus 2, respectively.

Conclusions: The broad and variable content of Medicaid-related tweets necessitates automatic categorization to identify topic-relevant posts. Our proposed system presents a feasible solution for automatic categorization and can be deployed and generalized for health service programs other than Medicaid. Annotated data and methods are available for future studies.

(J Med Internet Res 2021;23(5):e26616) doi:10.2196/26616

XSL•FO
RenderX

KEYWORDS

natural language processing; machine learning; Twitter; infodemiology; infoveillance; Twitter; social media; Medicaid; consumer feedback

Introduction

Consumers' perspectives and feedback are crucial for improving products or services. Over the last two decades, widespread adoption and use of the internet has led to its use as a major platform for collecting targeted consumer feedback. Businesses often allow consumers to rate specific products and services and provide detailed comments or reviews, and this has become a key feature of e-commerce platforms. For example, consumer-generated reviews and ratings of products play an important role in the differentiation on Amazon's e-commerce site, which currently has a global presence [1,2]. There are also companies, such as Yelp, that focus specifically on crowdsourcing consumer feedback [3-6]. Similarly, as social media has become the primary platform of communication for many people, many companies have started maintaining and communicating via social media accounts, often enabling direct communications, both private and public, with consumers. Not only do consumers provide comments or seek assistance through these social media accounts, they also often engage in discussions about products or services within their own social networks. Consequently, such consumer-generated chatter is often used to assess perceptions about specific topics, which may range from products or services to social programs, legislation, and politics.

Social media is a rich resource for obtaining perspectives on public health, as it enables the collection of large amounts of data directly and in real time. It is commonly used for sentiment analysis—a field of study that analyzes opinions, sentiments, attitudes, and emotions from written language. Sentiment analysis research involving social media data has covered a wide range of topics, events, individuals, issues, services, products, and organizations [7,8]. However, the use of social media has not been limited to sentiment analysis in open domains. In recent years, research within the broader medical domain has embraced social media, and it is currently being used for conducting real-time public health surveillance, including for topics such as influenza surveillance, pharmacovigilance, and toxicovigilance [9-11]. Meanwhile, similar to corporate businesses in the United States, health service providers such as local health departments and hospitals have also started adopting social media specifically as a consumer-facing communication channel [12,13]. Prior studies in this area have investigated how social media data linked to such health services accounts reflect the consumers' perspectives about them. The simplest studies have focused on using structured or numeric information, such as likes or ratings, associated with the accounts belonging to hospitals or nursing homes, and these metrics have been compared with traditional quality reports and ratings [14-16]. Building on the advances in open-domain natural language processing (NLP), some studies within the broader health domain have attempted to use unstructured data, including postings related to patients' experiences about hospitals, to infer consumer sentiments [17,18] or extract topics that summarize content [19].

Extracting knowledge from social media data is notoriously difficult for NLP methods because of factors such as the presence of misspellings, colloquial expressions, lack of context, and noise. These problems are exacerbated for health-related data because of the complexities of domain-specific terminologies, the lack of expert knowledge among common social media users, and the uniqueness of health-related topics. Consequently, there is considerably less research using the free-text data on social media for health-related tasks. Past studies closely related to ours have focused on analyzing sentiments toward attributes of health insurance plans [20] and social media users' responses to public announcements about health policies [21]. However, to the best of our knowledge, there has been no near real-time automatic system that provides comprehensive data collection and analysis on social media chatter about health services and insurance coverage provided by large public insurers such as Medicaid and Medicare.

Nevertheless, such a system is essential for analyzing the public's perspective toward public insurers and their governance and policies from social media. For example, the customers using Medicaid might provide their feedback based on their experiences or even engage in discussion of their experiences, which they might not have a chance or even willing to reveal to the Medicaid providers' customer service representatives. Analyzing such chatter could provide researchers and policy makers with information complementary to traditional customer feedback channels and possibly help improve the services and related policies. However, chatter associated with an entity such as Medicaid contains discussions about politics and legislation; academic research, statistics, and factual information; consumer feedback; and so on. Chatter related to politics can be very different in terms of content, compared with chatter related to consumer feedback. Thus, properly categorizing these tweets based on content is crucial for providing accurate information. Furthermore, sentiment may also have different meanings for these 2 broad categories of chatter—negative sentiment in political chatter may represent a user's emotions associated with a political decision about the health service (eg, changes in policies related to insurance coverage or covered benefits within Medicare or Medicaid) rather than the service itself.

Therefore, there is a need to identify and categorize the content of the chatter before it can be used for targeted analyses. A good categorization scheme can not only help in bringing forth good analysis but can also help to avoid contaminating the chatter with irrelevant content. To achieve this and effectively use social media big data, automatic classification and analysis systems based on machine learning methods are required. This, together with the promise of social media data and the lack of past research in this specialized area, served as the primary motivation for this study. We chose Medicaid as our target health service because it is the single largest public insurance

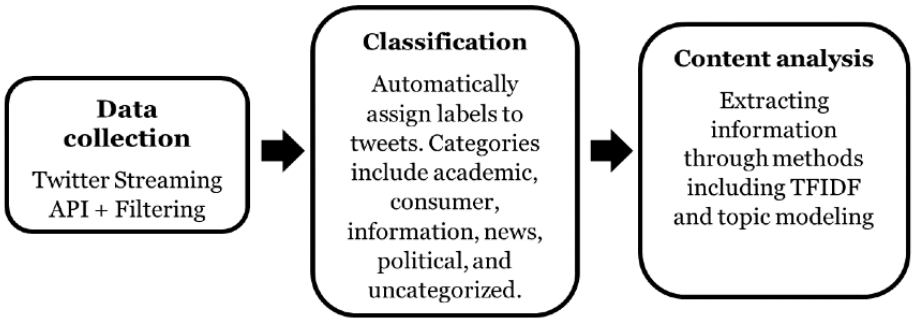
XSL•FO
RenderX

program in the country [22] and contains large volumes of related chatter on social media.

The specific objectives of this paper are as follows:

- To assess if a social media platform, specifically Twitter, contains sufficient volumes of chatter about health services so that it can be used to conduct large-scale analyses, using Medicaid as our target service
- To develop and discuss a data-centric system involving NLP and machine learning to automatically collect, categorize, and analyze Twitter chatter associated with Medicaid, as shown in Figure 1
- To describe the manual annotation of a Twitter-Medicaid data set and its composition
- To describe supervised classification strategies for automatically classifying Medicaid-related tweets into broad categories and evaluating the performances of several machine learning models, with particular emphasis on tweets that potentially represent consumer feedback
- To conduct postclassification content analyses to verify the potential utility of our data-centric system

Figure 1. Workflow of the natural language processing system for automatic data collection, classification, and content analysis of the Medicaid chatter on Twitter. API: application programming interface; TFIDF: term frequency-inverse document frequency.



The main contributions of this paper are as follows:

- We present the methods and results of collecting Medicaid-related Twitter data, analyzing a sample of the data manually, and developing an annotation guideline suitable for preparing a large data set for training classification algorithms.
- We present details of automatic supervised classification experiments, including methods, results, and evaluations, and provide suggestions on how to further improve the performance.
- We discuss the postclassification analyses of the collected data, including data distribution and content analyses.
- We make the NLP and machine learning scripts in this study publicly available, along with the labeled training data set and a larger set of unlabeled Medicaid-related data.

Methods

Data Collection

To develop our models for analyses of Twitter data related to Medicaid, we collected 2 sets of publicly available data from the network, which we labeled Corpus 1 and Corpus 2. Corpus 1 contains tweets mentioning the term "Medicaid," or Medicaid agency (MA) and managed care organization (MCO, an organization that provides Medicaid-related health services under contracts from the agency) names that are branded and thus easily distinguishable on Twitter (eg, *Medi-cal*: California's Medicaid program, and *TennCare*, Tennessee's Medicaid program). These tweets were collected via Twitter's public streaming application programming interface from May 1, 2018, to October 31, 2019, and were limited to only English tweets. It has been reported that misspellings appear frequently on social media platforms [23], particularly Twitter; hence, we used an automatic spelling variant generator to generate common misspellings for "medicaid" and used them to capture tweets referring "medicaid" as one of the misspellings [24]. This can increase the retrieval rate and increase the volume of streaming data. The full list of keywords, including misspellings, is shown in Table S1 in Multimedia Appendix 1. We then identified and removed tweets whose contents were not directly related to Medicaid and tweets with repeated or duplicated contents (eg, fundraising or political campaigns). To focus on tweets expressing personal opinions, we removed retweets, which were deemed as duplicates of the original tweets. The final data set consisted of 628,411 tweets for Corpus 1.

Although most of the chatter regarding Medicaid posted by consumers only included the term "medicaid" (or its variants), some directly tagged or mentioned relevant Twitter handles associated with MAs or the MCOs (eg, "@organization_name"). Corpus 2 is composed of such tweets, and the MA and MCO Twitter handles were identified in a previous study [25]. The full list of the handles used in data collection is presented in Table S2 in Multimedia Appendix 1. These tweets were retrieved by targeted searching (eg, "to:organization_name") on Twitter. These tweets were posted between December 12, 2008, and the time of search (January 9, 2020). We filtered the tweets using the same approaches used for Corpus 1. Overall, there were 27,337 tweets in the corpus. Additional notes regarding our data are provided in Multimedia Appendix 1.

Tweet Contents and Manual Annotations

To better understand the contents of the tweets posted by users and to develop methods to automatically characterize the posts, we first performed manual inspections of the contents of the posts and identified commonly occurring themes. We used the grounded theory approach to conduct a thorough analysis [26]. We analyzed a random sample of tweets to identify recurring topics and then grouped the topics into broader categories and themes. The analysis was conducted by multiple authors of this

XSL•FO
RenderX

paper, and the topics discovered initially were discussed. The discovered topics were either merged into broader themes (eg, combining *information* and *outreach*), discarded from our consideration (eg, for topics that were observed rarely or only once), or split into multiple themes (eg, splitting of *information* tweets into *academic*, *information*/*outreach*, and *news*). Following iterative discussions and finalization by the domain expert authors of this paper (JMZ and DG), we classified tweet contents into 5 broad categories: (1) academic, (2) consumer feedback, (3) information/outreach, (4) news, and (5) political opinion/advocacy. Tweets that could not be categorized as any of these were labeled as *other*. The descriptions of these classes are as follows:

- *Academic* (*academic*)—tweets related to research on Medicaid. These included tweets by persons or organizations with academic affiliations or think tanks that expressed the perspective from the affiliated organizations or any tweet relating to education, scholarship, and thought, including (links to) journal publications and reports.
- *Consumer feedback* (*consumer*): These included tweets related to consumers' experiences or questions related to Medicaid services, coverage, benefits, or health issues. The tweets were typically from Medicaid consumers or family members of consumers and also included discussions with others.
- *Information/outreach* (*information*): These included tweets directed at consumers and beneficiaries of Medicaid to convey information including agency services, programs, events, enrollment, eligibility criteria, etc. Tweets containing information about general health or public health reminders were also included.
- *News* (*news*): These included news and announcements, including any tweets from a news agency or organization. Tweets that explicitly expressed political opinions and tweets from Medicaid agencies or plans were excluded.
- *Political opinion/advocacy* (*political*): These included comments, personal opinions, and feedback about politics related to Medicaid.
- *Other* (*other*): These included tweets that were not relevant, typically the noise that is not captured by the initial screening.

Following the establishment of the desired categories and the development of annotation guidelines by JMZ, 2 trained annotators performed the first round of annotations (for the data in Corpus 1) in multiple iterations, developed annotation guidelines, and resolved ambiguities via discussions. Following the completion of this round of annotations, the annotation disagreements were resolved by AS and WH. We found the class distribution to be very imbalanced, with most of the tweets annotated as *news*, *political*, and *other*, whereas only a small portion were in *academic*, *consumer*, and *information* categories (Table 1). Examples of each category are provided in Table S3 in Multimedia Appendix 1. To understand how this imbalanced distribution affected the classifier performances on the smaller classes, particularly the consumer class, we performed preliminary automatic classification experiments using 3 classifiers: naïve Bayes (NB), support vector machine (SVM), and random forest (RF). We split the data into training (5795/7244, 80%) and validation (1449/7244, 20%) sets and found the best performance on consumer feedback to be low for all the classifiers, with the best F_1 score=0.3 (SVM). Tweets belonging to the consumer feedback class were of particular importance to our overarching project objectives, so we devised 2 strategies for improving performance for this class: the first involved additional annotations of targeted tweets from the same data set and the second focused on collecting an additional data set (Corpus 2, as described earlier).

Table 1. Distribution (counts and percentages) of annotated data in the first round of annotations (rows 2 and 3) and the final data sets (Corpus 1 for rows 4 and 6; Corpus 2 for rows 5 and 7).

Data set	Academic, n (%)	Consumer, n (%)	Information, n (%)	News, n (%)	Political, n (%)	Other, n (%)	Total, n (%)
Training set (first round)	61 (1.05)	158 (2.73)	198 (3.42)	1288 (22.23)	3613 (62.34)	477 (8.23)	5795 (100)
Validation set (first round)	35 (2.42)	37 (2.55)	49 (3.38)	317 (21.88)	897 (61.90)	114 (7.86)	1449 (100)
Training set (Corpus 1)	83 (1.23)	355 (5.27)	429 (6.37)	1299 (19.30)	3710 (55.13)	854 (12.69)	6730 (100)
Training set (Corpus 2)	9 (0.75)	709 (59.08)	94 (7.83)	40 (3.33)	10 (0.83)	338 (28.17)	1200 (100)
Test set (Corpus 1)	20 (2)	46 (4.60)	49 (4.90)	199 (19.90)	603 (60.30)	83 (8.30)	1000 (100)
Test set (Corpus 2)	6 (0.60)	579 (57.90)	80 (8.00)	21 (2.10)	6 (0.60)	308 (30.80)	1000 (100)
Total	153 (1.34)	1726 (15.17)	701 (6.16)	1876 (16.49)	5226 (45.93)	1697 (14.91)	11,379 (100)

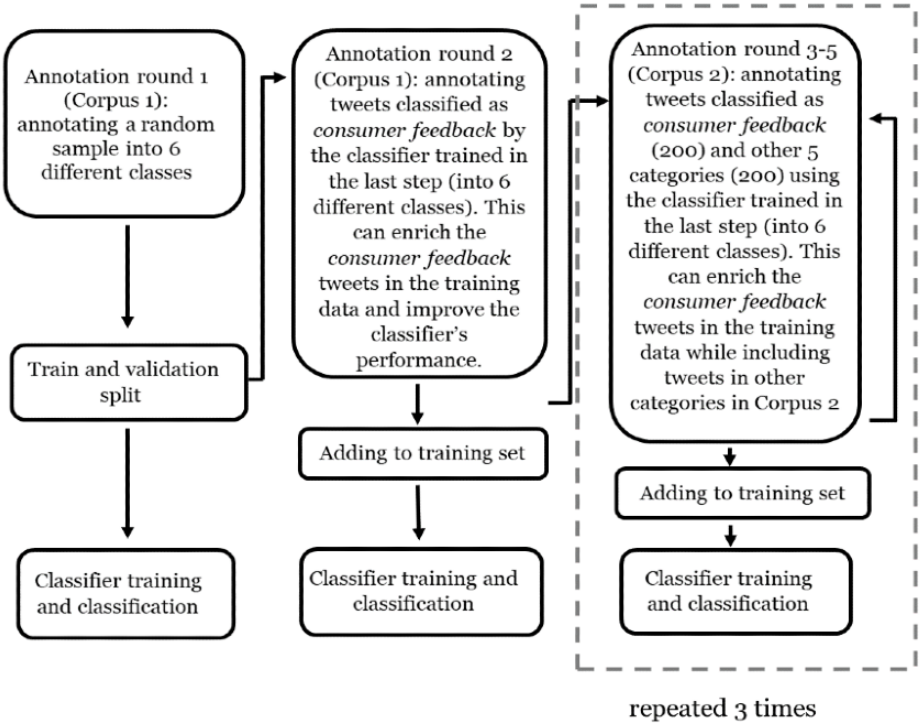
For the first strategy, we conducted another round of annotation of tweets from Corpus 1 to increase the number of tweets for the consumer class. Owing to the very low number of consumer class tweets in the original data set, we realized that it would not be feasible to annotate sufficient numbers of these tweets by drawing random samples because of budgetary and other constraints. Therefore, rather than randomly drawing tweets for the next round of annotations, which would again lead to finding a small number of tweets belonging to the consumer feedback category, we attempted to artificially increase the number of tweets for this category. We achieved this by running our above-described weak classifier on a larger set of unlabeled tweets and only picking tweets classified as consumer feedback by the SVM classifier. This significantly increased the number of consumer feedback tweets in the data to be annotated. The

XSL•FO
RenderX

new set of annotated data was then added to the training set, and the data distribution is presented in Table 1.

We followed the same annotation strategy for Corpus 2 (ie, annotating tweets classified as consumer feedback by the classifier trained on the previously annotated data), but this time, we also annotated equal amount of nonconsumer tweets. This is because Corpus 2 is rich in consumer feedback tweets, and we would also like to include tweets in other categories to improve performance. An outline of the overall annotation process is presented in Figure 2. Although we tried to decrease the class imbalance in the training sets of the 2 corpora, to ensure that our evaluations represented the classifier performances on real-world distributions of the data, we did not artificially balance the validation set. We also annotated the test set randomly generated from the 2 corpora, 1000 tweets each, so they would reflect the data composition of the original corpora, allowing us to evaluate how the classifier would perform when deployed for streaming data.

Figure 2. Flowchart for the entire annotation process (for the training and validation set) involving multiple rounds.



Classification

We experimented with 5 traditional classification algorithms, including Gaussian NB [27,28], SVM [29,30], RF [31], k-nearest neighbor (KNN) [28], and shallow neural network (NN), and 2 advanced classification algorithms, bidirectional long short-term memory (BLSTM) [32,33] and bidirectional encoder representations from transformers (BERT) [34,35]. Although the origin and distributions of tweets in the 2 corpora were different, we decided to combine them as our previous research suggests that multicorpus training, or distant supervision, leads to performance improvements for social media text classification [36]. The feature extraction and classification training for traditional classifiers was done using the "Scikit-learn" package in Python [37], the BLSTM classification was implemented using package "Keras" in Python [38], and the BERT classification approach was implemented using package "simpletransformers," which is based on the package "transformers" [39]. The performance on the validation set and the test set from Corpus 1 and Corpus 2 are shown in Table 2.

XSL•FO
RenderX

Table 2. Classification performances of the classifiers on the test sets of Corpus 1 and Corpus 2.

Data set and classification algorithm	F_1 score (0.XX)[a]						Percentage accuracy (95% CI)
	Academic	Consumer (95% CI)	Information	News	Political	Other	
Validation set							
NB[b]	11	17 (11-24)	24	55	70	34	55.0 (52.4-57.6)
SVM[c]	0	*53 (38-66)* [d]	26	70	87	43	77.4 (75.2-79.5)
RF[e]	5	43 (26-58)	27	74	87	48	78.7 (76.6-80.7)
KNN[f]	5	24 (12-37)	11	55	65	26	51.4 (48.9-54.0)
NN[g]	31	34 (21-46)	32	72	86	46	75.2 (72.9-77.4)
BLSTM[h]	27	38 (25-51)	42	74	88	53	78.9 (76.8-81.0)
BERT[i]	54	*61 (48-72)*	64	82	*92*	67	*85.2 (83.3-87.0)*
Test set (Corpus 1)							
NB	12	23 (16-31)	20	53	71	21	53.5 (50.4-56.6)
SVM	0	38 (24-51)	14	71	83	19	73.0 (70.2-75.7)
RF	0	24 (10-37)	21	75	84	24	74.6 (71.9-77.2)
KNN	0	20 (9-32)	15	47	66	26	49.0 (45.9-52.1)
NN	25	*44 (31-56)*	33	70	84	32	71.8 (69.0-74.6)
BLSTM	22	33 (19-45)	20	71	84	30	73.1 (70.4-75.8)
BERT	72	*58 (45-70)*	58	80	*89*	51	*81.7 (79.3-84.0)*
Test set (Corpus 2)							
NB	0	72 (69-75)	30	11	3	21	47.3 (44.2-50.3)
SVM	0	76 (73-78)	2	21	7	18	56.4 (53.3-59.4)
RF	0	*82 (80-84)*	7	16	11	66	69.4 (66.6-72.3)
KNN	0	38 (33-42)	0	7	0	50	42.2 (39.1-45.3)
NN	0	79 (76-82)	40	24	5	66	66.0 (63.0-69.0)
BLSTM	0	81 (79-84)	34	21	4	55	67.3 (64.4-70.2)
BERT	50	*90 (89-92)*	49	37	21	79	*80.7 (78.2-83.1)*

[a]The number represents the first two decimal points. For example, the F1 score for SVM on Consumer is 0.53 with 95% CI 0.38-0.66.

[b]NB: naïve Bayes.

[c]SVM: support vector machine.

[d]The best scores are highlighted in italics.

[e]RF: random forest.

[f]KNN: k-nearest neighbor.

[g]NN: shallow neural network.

[h]BLSTM: bidirectional long short-term memory.

[i]BERT: bidirectional encoder representations from transformers.

The tweets were preprocessed by lowercasing and anonymizing the URLs and user names. For the traditional classifiers, the non-English characters were further removed (keeping underline), and each word was stemmed by the Porter stemmer. The features were the unnormalized counts of the 3000 most frequent n-grams (contiguous sequences of words with n ranging from 1 to 3, with 1380 unigrams, 1296 bigrams, and 324 trigrams). We also introduced a "word cluster" feature, which are clusters or generalized representations of semantically similar words or phrases learned from Twitter chatter [10,40]. The word clusters were represented as bag-of-word vectors, and the feature space consisted of 972 word clusters. We used the Twitter word clusters, "50mpaths2," provided by Owoputi et al [41]. For the advanced classifiers, each word or character sequence was replaced with a dense vector, and the vectors were then fed into the relevant algorithms for training.

We performed hyperparameter tuning using the validation set to improve the classification task on the imbalanced data set. Specifically, we focused on improving the F_1 score for consumer

XSL•FO
RenderX

feedback. For traditional classifiers, we optimized the number of nearest neighbors for KNN, the number of estimators (trees) for RF, and the c parameter and weights for SVM. We also experimented with oversampling using the synthetic minority oversampling technique, but the performance was not improved (provided in Table S4 in Multimedia Appendix 1). The optimal hyperparameters for the traditional classifiers are listed in Table S5 in Multimedia Appendix 1. We used the Twitter GloVe word embeddings for the BLSTM [42] classifier, where each word was converted to a 200-dimensional vector. BLSTM was then trained with 40 epochs and dropout regularization, and the best model was selected based on the accuracy of the validation data. We chose RoBERTa-large for the BERT algorithms [35], trained with 3 epochs. The technical details are provided in Table S5 in Multimedia Appendix 1.

Postclassification Analyses

To assess the utility of our classification approaches and gain an understanding of the data, we used the best-performing classifier (the classifier based on BERT) to label all collected unlabeled data and compute the data distribution. We then performed content analysis using the term frequency-inverse document frequency (TFIDF) method [43], focusing on the tweets in Corpus 1 that contained the term “medicaid” and its misspellings, and using the latent Dirichlet allocation (LDA) for topic modeling [44], focusing on consumer feedback tweets. Our first intent was to qualitatively assess whether the classifier was capable of distinguishing tweets based on contents that were manually verifiable. Second, we wanted to obtain a basic understanding of the content of each category by identifying the top rated TFIDF words. The TFIDF method adjusts the term frequencies with inverse document frequency so that the high-frequency terms unique in one category would rank higher than the high-frequency words that are common across the categories. This helps identify important terms unique to the target category. Our third objective was to summarize consumer feedback chatter using LDA topic modeling, going beyond the TFIDF method. For all content analyses, the text was first preprocessed by lowercasing and removing URLs, user names, non-English characters (keeping underline and hyphen), stopwords, and any word with less than 4 characters. For LDA topic modeling, we experimented with different hyperparameters (number of topics=5, 10, 20, 50, and 100) and selected the model with the highest coherence score.

Results

Annotation and Class Distributions in Test Sets

We annotated a total of 9179 tweets from Corpus 1 and 2200 tweets from Corpus 2. We obtained substantial interannotator agreement (Cohen κ=0.734) [45,46] over 892 double-annotated tweets. The test data sets were randomly selected from the corpora, and therefore, they can be considered a sample of the collected data. For Corpus 1, the test data contained 1000 tweets, among which political discussion was the dominant class (603/1000, 60.30%), followed by news (199/1000, 19.90%), whereas consumer feedback comprised 4.60% (46/1000) of the tweets. In contrast, consumer feedback comprised 57.90% (579/1000) of the tweets in Corpus 2, and 30.80% (308/1000) of the tweets could not be categorized, most of which were part of conversations and could not be understood without full context.

Classification Results

The F_1 scores for each class and the accuracies of the classifiers on the validation set and the test sets are presented in Table 2, including CIs estimated using bootstrapping, whereas the precisions and recalls are given in Table S6 of Multimedia Appendix 1. For the validation and test sets from Corpus 1, the classifiers showed high performance for political discussion, but relatively low performance for consumer feedback. This was expected based on the large imbalance described earlier. Among all the traditional classifiers tested, SVM performed the best on the validation set, with an F_1 score of 0.53 on the consumer feedback. However, the F_1 score on the consumer feedback on the test set from Corpus 1 was only 0.38. In contrast, we found that the BERT classifier had the highest F_1 scores on consumer feedback for both the validation set (0.61) and the test set from Corpus 1 (0.58).

For the test set from Corpus 2, most of the classifiers performed well on the consumer feedback. Among the traditional classifiers, RF performed the best, with an F_1 score of 0.82 on consumer feedback. On the other hand, BERT still performed the best, with a consumer feedback F_1 score of 0.90.

As the BERT classifier performed the best in terms of accuracy and the consumer feedback F_1 score on the validation set and the 2 test sets, we used the BERT classification for postclassification analysis.

Error Analysis

We conducted a brief analysis of the errors made by the BERT-based classifier. We first calculated the confusion matrix for both test sets (Table 3). In Table 4, we provide examples of the most frequent classification errors, omitting unnecessary details. For Corpus 1, we highlighted that the classifier frequently misclassified political tweets as news or consumer feedback, and vice versa. This is not surprising because users sometimes commented on and discussed politics with personal experience, and some news content was related to opinions about the policy. We also highlighted that the uncategorized tweets, whose content is often not directly related to Medicaid or lack of information, are frequently misclassified as consumer feedback or political. The confusion between consumer feedback and political or uncategorized tweets, along with the low volume of consumer feedback, contributes to the low performance of consumer feedback. We also observed that some news tweets were confused with the information tweets because information is frequently spreading as news or blog articles.

XSL•FO
RenderX

Table 3. The BERT classifier's confusion matrix on the test set

Data set and true value	Predicted value					
	Academic	Consumer	Information	News	Political	Other
Test set (Corpus 1)						
Academic	13	0	1	4	2	0
Consumer	0	26	0	0	18	2
Information	0	0	27	9	10	3
News	1	0	9	169	17	3
Political	2	5	3	39	549	5
Other	0	12	4	4	30	33
Test set (Corpus 2)						
Academic	3	0	0	1	2	0
Consumer	1	512	1	2	26	37
Information	1	5	33	15	1	25
News	0	0	5	11	3	2
Political	0	0	0	1	5	0
Other	1	36	15	7	6	243

Table 4. Examples of misclassified tweets by the BERT classifier on Corpus 1 and Corpus 2.

Data set, Tweets	True class (prediction)	Comments
Test set (Corpus 1)		
I need this government shutdown to end because no one is going to call me to set up my Medicaid while it's shutdown	Political (consumer)	Discussion about politics with personal experience
"This is just cruelty and exclusion": Amid Trump's attack on poor, one million fewer kids receiving Medicaid and CHIP–Raw Story <URL>	Political (news)	Opinion on Medicaid policy presented as a news title
<USERNAME> So do I! But I totally understand why some people really hate it. And yes... lack of Medicaid providers is a problem everywhere (I do accept it, but only have a mobile practice). Maybe contact your local health department and ask!	Consumer (political)	Customer's discussion about Medicaid services. It may have been misclassified because of similarity to political discussion regarding Medicaid
Thanks to <USERNAME> for this story about the bill ... Ohio leaving some military families with special needs children waiting for answers <URL>	News (political)	News about Medicaid policy reformation bill
States that been successful in lowering substance use disorder rates have increased access to medicaid & private insurance, and to MAT and naloxone. Thank you, NYT Editorial Board @NYTOpinion. <URL>	News (information)	News about information related to Medicaid
3 Ways to increase Missouri Medicaid EMOMED Reimbursement <URL>	Information (news)	Information for Medicaid beneficiaries, presented as a blog article
The Medicaid office I'm going to tomorrow opens at 7:30 am. I won't be there that early, but ugh.	Other (consumer)	Uncategorized because it is not about experience or question, but content indicates the user to be a customer
<USERNAME> I hope someone will ask him "What's the difference between Medicaid and Medicare?"	Other (political)	Uncategorized because of lack of related content but is similar to political discussion
Test set (Corpus 2)		
<organization_name> poorly worded	Consumer (other)	Most likely comment on customer service but hard to pick up by algorithm
<organization_name> My pleasure!	Other (consumer)	Classified as others because of lack of information, but the algorithm might recognize that it could be a conversation between a customer and a customer representative

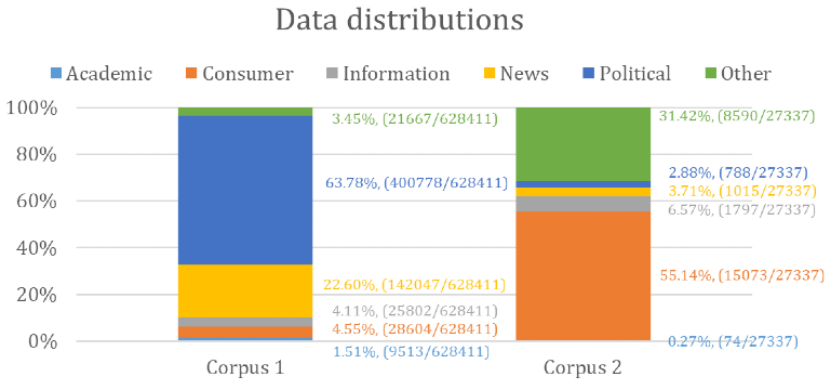
XSL•FO
RenderX

For Corpus 2, the dominant classes were consumer feedback and uncategorized tweets, and they were most frequently misclassified as each other. We suspect they were misclassified because tweets sometimes lacked context, making their meanings ambiguous and difficult for the machine to understand. For example, the tweet "<organization_name> poorly worded," though ambiguous, might be understood as that some document for the customer or the customer service representative's expression was poorly worded, and thus, we categorized it as consumer feedback. However, the machine learning algorithms were not capable of deciphering such implicit contexts: "poorly worded" is usually associated with a feedback and, in tweets directed to the agency's handle, it is likely related to customer service. Similarly, the tweet "<organization_name> My pleasure!," may belong to a conversation between a customer and a representative, but the lack of information renders it to the other class. However, machine learning could not capture this understanding.

Postclassification Analyses: Data Distribution

We applied the best-performing classifier (BERT) to label both corpora. The obtained class distributions are shown in Figure 3. We found that the majority of tweets in Corpus 1 were news (142047/628411, 22.60%) and political discussion (400778/628411, 63.78%), whereas consumer feedback accounted for only 4.55% (28604/628411), consistent with the data distribution of the test set of Corpus 1. The data distribution indicates that this corpus is suitable for analyzing chatter regarding political discussion or news. For Corpus 2, the majority of the tweets were labeled as consumer feedback (15073/27337, 55.14%) and uncategorized (8590/27337, 31.42%), which is also consistent with the data distribution in the test set.

Figure 3. Postclassification class distributions among 2 corpora, as per the automatically classified tweets.



Postclassification Analyses: Content of Each Class in Corpus 1

We now briefly summarize the findings from content analyses of the tweets in Corpus 1 that contain the terms associated with "medicaid" to understand, from a high-level perspective, the contents within each category. The 10 highest ranking bigrams and trigrams detected by the TFIDF method are listed in Table S7 of Multimedia Appendix 1 [43]. Not surprisingly, the academic tweets are dominated by terms starting with "study..." and terms indicating research finding. Similarly, the information tweets contain terms related to "service," "care,"... etc, consistent with information outreach. For the news tweets, we found that many tweets were about news on medicaid work requirements in Kentucky and Arkansas (blocked by the federal judge on March 27, 2019). In addition, "social security" and "Trump..." are also highly ranked among the news and political classes. For the tweets belonging to the consumer feedback, some of the high-ranking terms were shared with other classes (eg, "... insurance," "social security," or "... care"), whereas some were specific to this class ("make much" or "doesn cover") and potentially indicated comments about Medicaid income cap and coverage.

We did not know the compositions of the 2 data sets that we had collected a priori. Thus, the results of our classification experiments provided us with very important knowledge about which type of Twitter data to use when conducting targeted studies about Medicaid or health services in general. For example, when studying consumer feedback, it is best to use data from Corpus 2 (ie, tweets containing Twitter handles of the MA or MCO); for studying public perceptions of political decisions, Corpus 1 would be more useful. Detailed content analyses of the tweets in each category, such as their temporal and geolocation-specific distributions, are likely to reveal more relevant information. However, such analyses are outside the scope of this study, and we plan to build on the NLP system described in this paper to conduct more thorough content analyses in the future.

Postclassification Analyses: LDA Topic Modeling on Consumer Feedback

We found that the model with 20 topics achieved the highest coherence score. The top 20 words in each topic are listed in Multimedia Appendix 1, Table S8. We now summarize the main findings based on these top words, with example top words provided in parentheses. We deduced that this chatter contains discussion related to (1) applying for Medicaid, either for oneself or even family members (eg, *deny*, *apply*, *family*, and *child*); (2) Medicaid coverage for dentists, specialists, prescription medications, emergency department visits (eg, *cover*, *dentist*, *therapist*, *prescription*, *medication*, and *emergency*); (3) interacting with customer representatives, especially through phone (eg, *call*, *phone*, *hour*, *hold*, *tell*, and *wait*); (4) hospital-related bills (eg, *hospital*, *bill*, and *copay*); and (5) comparing different insurance plans (*switch*, *insurance*, *private*,

XSL•FO
RenderX

and *plan*). The list of topics can be a guide to further categorize the consumer feedback chatter, which could lead to a more detailed analysis and even provide recommendations on how to further improve the Medicaid program. A more in-depth analysis is left for future work.

Discussion

Principal Findings

As many classification errors occur because the tweets lie in the boundary between 2 classes, we note that a multilabel classification scheme might improve the performance [47]. However, in the experiments conducted earlier in this project, we found that the multilabel scheme only improved the classification performance by a small margin, while making the annotation process more difficult. Thus, we focused on the single-label classification scheme in this work, leaving the development of multilabel models to future work.

In addition to multilabel classification models, the classification error might also be remedied by creating new categories for tweets lying at the boundary of current categories. For example, we can further divide the political discussion into 2 categories: discussion of policy without personal experiences or experiences from friends or relatives and discussion of policy with experiences as supporting evidence. The classification performance may be further improved by including more user profile information. For example, we can include features such as whether the account belongs to a news agency or if the user is affiliated with an academic organization or think tanks, which could improve the classification performance on the news class or the academic class. As the 2 corpora have very different distributions, developing a corpus-specific classifier might further improve the performance.

Although our content analysis is limited to the high-ranking TFIDF terms and LDA topic modeling on consumer feedback, additional analyses could include topic modeling of other chatter [44] or sentiment analysis to understand sentiments toward Medicaid in general or to specific aspects of the Medicaid program [8]. The manual analysis of selected samples can deepen the understanding of these topics and potentially generate recommendations toward policy change. We also note that content analysis can not only help researchers further understand the Medicaid chatter but it can also improve the classification performance in reverse.

Limitations

This analysis has limitations related to the quality of Twitter data, which contain high volumes of noise that may affect the accuracy and generalizability of our content analyses and annotation guidelines. In addition, Twitter users may not be representative of Medicaid enrollees. Older age groups tend to be underrepresented among Twitter users [48], and more vulnerable populations who rely on Medicaid may not use this platform to discuss their health coverage.

Conclusions

We have developed a social media mining system, involving NLP and machine learning, for continuously collecting and categorizing Twitter chatter about the Medicaid program. Our study demonstrates that it is possible to collect data about large, complex health services and coverage programs such as Medicaid using Twitter to obtain near real-time knowledge about consumer perceptions and opinions. The automatic classification of streaming data is crucial, specifically for smaller classes, such as consumer feedback, for studying targeted topics.

Our analysis can inform public health researchers on how to use public discussions about health programs and services, such as Medicaid. Similarly, our system can be deployed by research groups or Medicaid agencies for continuous, ongoing research on the evolution of public opinions on social media (eg, the impact of certain policy changes or rulings). We also note that although this work focuses on Medicaid, our methods and open source code can readily be applied to other health services. Annotated data and methods are available for future studies [49].

Acknowledgments

This study was conducted with partial support from the Robert Wood Johnson Foundation (#76158, JMZ and DG).

Authors' Contributions

YY conducted and directed the machine learning experiments, evaluations, and data analyses with assistance from MAA, AS, and WH. YY, AS, MAA, and WH contributed to data collection, annotation, and analyses. JMZ and DG provided their expertise in preparing the annotation guidelines and categories and helped formulate the overarching objectives of the project. AS, JMZ, and DG supervised various aspects of the study. YY drafted the manuscript, and all the authors contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials, including tables and additional information.
[DOCX File , 51 KB - jmir_v23i5e26616_app1.docx]

References

XSL•FO
RenderX

1. Chen PS, Wu S, Yoon J. The impact of online recommendations and consumer feedback on sales. In: Proceedings of the International Conference on Information Systems, ICIS 2004. 2004 Presented at: International Conference on Information Systems, ICIS 2004; December 12-15, 2004; Washington, DC, USA URL: https://aisel.aisnet.org/icis2004/58/
2. 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]
3. Hu M, Liu B. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004 Presented at: KDD04: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2004; Seattle WA USA p. 168-177. [doi: 10.1145/1014052.1014073]
4. Akay A, Dragomir A, Erlandsson B. A Novel Data-Mining Approach Leveraging Social Media to Monitor Consumer Opinion of Sitagliptin. IEEE J. Biomed. Health Inform 2015 Jan;19(1):389-396. [doi: 10.1109/jbhi.2013.2295834]
5. Lim Y, Van Der Heide B. Evaluating the wisdom of strangers: the perceived credibility of online consumer reviews on yelp. J Comput-Mediat Comm 2014 Aug 25;20(1):67-82. [doi: 10.1111/jcc4.12093]
6. Luca M. Reviews, reputation, and revenue: the case of Yelp.com. Harvard Business School NOM Unit Working Paper 2016. [doi: 10.2139/ssrn.1928601]
7. Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowl Inf Syst 2018 Jul 4;60(2):617-663. [doi: 10.1007/s10115-018-1236-4]
8. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 2012 May 23;5(1):1-167. [doi: 10.2200/s00416ed1v01y201204hlt016]
9. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013 Dec 9;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
10. 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 9;39(3):231-240 [FREE Full text] [doi: 10.1007/s40264-015-0379-4] [Medline: 26748505]
11. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith K, Gonzalez G. Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions. AMIA Annu Symp Proc 2014:924-933 [FREE Full text] [Medline: 25954400]
12. Griffis HM, Kilaru AS, Werner RM, Asch DA, Hershey JC, Hill S, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264 [FREE Full text] [doi: 10.2196/jmir.3758] [Medline: 25431831]
13. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013 Sep;103(9):1700-1707. [doi: 10.2105/ajph.2012.301166]
14. Glover M, Khalilzadeh O, Choy G, Prabhakar AM, Pandharipande PV, Gazelle GS. Hospital Evaluations by Social Media: A Comparative Analysis of Facebook Ratings among Performance Outliers. J Gen Intern Med 2015 Oct 7;30(10):1440-1446 [FREE Full text] [doi: 10.1007/s11606-015-3236-3] [Medline: 25749881]
15. Campbell L, Li Y. Are Facebook user ratings associated with hospital cost, quality and patient satisfaction? A cross-sectional analysis of hospitals in New York State. BMJ Qual Saf 2018 Feb 09;27(2):119-129. [doi: 10.1136/bmjqs-2016-006291] [Medline: 28600452]
16. Hefele JG, Li Y, Campbell L, Barooah A, Wang J. Nursing home Facebook reviews: who has them, and how do they relate to other measures of quality and experience? BMJ Qual Saf 2018 Feb 05;27(2):130-139. [doi: 10.1136/bmjqs-2017-006492] [Medline: 28780533]
17. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Jun 13;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]
18. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and Analyzing Patient Experiences of Health Care From Social Media. JMIR Res Protoc 2015 Jul 02;4(3):e78 [FREE Full text] [doi: 10.2196/resprot.3433] [Medline: 26137885]
19. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
20. van den Broek-Altenburg EM, Atherly AJ. Using Social Media to Identify Consumers' Sentiments towards Attributes of Health Insurance during Enrollment Season. Applied Sciences 2019 May 17;9(10):2035. [doi: 10.3390/app9102035]
21. Hatchard JL, Neto JQ, Vasilakis C, Evans-Reeves KA. Tweeting about public health policy: social media response to the UK Government's announcement of a parliamentary vote on draft standardised packaging regulations. PLoS One 2019 Feb 26;14(2):e0211758 [FREE Full text] [doi: 10.1371/journal.pone.0211758] [Medline: 30807582]
22. Medicaid enrollment changes following the ACA. Medicaid and CHIP Payment and Access Commission. URL: https://www.macpac.gov/subtopic/medicaid-enrollment-changes-following-the-aca/ [accessed 2020-03-11]
23. Han B, Cook P, Baldwin T. Lexical normalization for social media text. ACM Trans Intell Syst Technol 2013 Jan;4(1):1-27. [doi: 10.1145/2414425.2414430]
24. Sarker A, Gonzalez-Hernandez G. An unsupervised and customizable misspelling generator for mining noisy health-related text sources. J Biomed Inform 2018 Dec;88:98-107 [FREE Full text] [doi: 10.1016/j.jbi.2018.11.007] [Medline: 30445220]

XSL•FO
RenderX

25. Zhu JM, Sarker A, Gollust S, Merchant R, Grande D. Characteristics of Twitter use by state medicaid programs in the United States: machine learning approach. J Med Internet Res 2020 Aug 17;22(8):e18401. [doi: 10.2196/18401] [Medline: 32804085]
26. Martin PY, Turner BA. Grounded theory and organizational research. J Appl Behav Sci 1986 Apr 01;22(2):141-157. [doi: 10.1177/002188638602200207]
27. Rish I. An empirical study of the naive Bayes classifier. In: Proceedings of IJCAI-01 workshop on Empirical Methods in AI. 2001 Presented at: IJCAI-01 workshop on Empirical Methods in AI; 2001; Sicily, Italy p. 41-46 URL: https://www.cc.gatech.edu/~isbell/reading/papers/Rish.pdf
28. Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 1967 Jan;13(1):21-27. [doi: 10.1109/tit.1967.1053964]
29. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011 Apr;2(3):1-27. [doi: 10.1145/1961189.1961199]
30. Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. Cambridge, MA: The MIT Press; 1999:61-74.
31. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995 Presented at: 3rd International Conference on Document Analysis and Recognition; August 14-16, 1995; Montreal, QC, Canada. [doi: 10.1109/icdar.1995.598994]
32. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
33. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673-2681. [doi: 10.1109/78.650093]
34. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. Computation and Language. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-04-12]
35. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: a robustly optimized bert pretraining approach. Computation and Language. 2019. URL: https://arxiv.org/abs/1907.11692 [accessed 2021-04-12]
36. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J Biomed Inform 2015 Feb;53:196-207 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.002] [Medline: 25451103]
37. 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 Nov:2825-2830 [FREE Full text] [doi: 10.5555/1953048.2078195]
38. Keras Computer Program. 2015. URL: https://github.com/keras-team/keras [accessed 2021-04-12]
39. Wolf T, Debut L, Sanh V. Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020 Presented at: Conference on Empirical Methods in Natural Language Processing: System Demonstrations; October, 2020; Online, EMNLP p. 38-45. [doi: 10.18653/v1/2020.emnlp-demos.6]
40. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
41. Owoputi O, O'Connor B, Dyer C, Gimpel K, Schneider N. Part-of-speech tagging for Twitter: Word clusters and other advances. Technical Report CMU-ML-12-107, Carnegie Mellon University 2012:- [FREE Full text]
42. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
43. Beel J, Gipp B, Langer S, Breitinger C. Research-paper recommender systems: a literature survey. Int J Digit Libr 2015 Jul 26;17(4):305-338. [doi: 10.1007/s00799-015-0156-0]
44. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
45. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
46. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [doi: 10.1037/h0026256] [Medline: 19673146]
47. Tsoumakas G, Katakis I. Multi-label classification: an overview. Int J Data Warehous Mining (IJDWM) 2007;3(3):1-13. [doi: 10.4018/jdwm.2007070101]
48. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2021-02-04]
49. Bitbucket. URL: https://yyang60@bitbucket.org/sarkerlab/medicaid-classification-script-and-data-for-public [accessed 2021-04-21]

Abbreviations

BERT: bidirectional encoder representations from transformers

XSL•FO
RenderX

BLSTM: bidirectional long short-term memory
KNN: k-nearest neighbor
LDA: latent Dirichlet allocation
MA: Medicaid agency
MCO: Managed Care Organization
NB: naïve Bayes
NLP: natural language processing
NN: shallow neural network
RF: random forest
SVM: support vector machine
TFIDF: term frequency-inverse document frequency

Edited by G Eysenbach; submitted 18.12.20; peer-reviewed by S Doan, K Acquaviva; comments to author 12.01.21; revised version received 08.02.21; accepted 03.04.21; published 03.05.21.

Please cite as:
Yang YC, Al-Garadi MA, Bremer W, Zhu JM, Grande D, Sarker A
Developing an Automatic System for Classifying Chatter About Health Services on Twitter: Case Study for Medicaid
J Med Internet Res 2021;23(5):e26616
URL: https://www.jmir.org/2021/5/e26616
doi:10.2196/26616
PMID:33938807

©Yuan-Chi Yang, Mohammed Ali Al-Garadi, Whitney Bremer, Jane M Zhu, David Grande, Abeed Sarker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2021. 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

Information-Seeking Patterns During the COVID-19 Pandemic Across the United States: Longitudinal Analysis of Google Trends Data

Tichakunda Mangono[1], BSc, MSc; Peter Smittenaar[1], BSc, MSc, PhD; Yael Caplan[1], BA; Vincent S Huang[1], BSc, PhD; Staci Sutermaster[1], BSc, MSc; Hannah Kemp[1], BA, MPA; Sema K Sgaier[1,2,3], BSc, MSc, PhD

[1]Surgo Ventures, Washington, DC, United States

[2]Department of Global Health & Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[3]Department of Global Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Tichakunda Mangono, BSc, MSc
Surgo Ventures
1701 Rhode Island Ave NW
Washington, DC, 20036
United States
Phone: 1 8579397670
Email: tichmangono@surgoventures.org

Abstract

Background: The COVID-19 pandemic has impacted people's lives at unprecedented speed and scale, including how they eat and work, what they are concerned about, how much they move, and how much they can earn. Traditional surveys in the area of public health can be expensive and time-consuming, and they can rapidly become outdated. The analysis of big data sets (such as electronic patient records and surveillance systems) is very complex. Google Trends is an alternative approach that has been used in the past to analyze health behaviors; however, most existing studies on COVID-19 using these data examine a single issue or a limited geographic area. This paper explores Google Trends as a proxy for what people are thinking, needing, and planning in real time across the United States.

Objective: We aimed to use Google Trends to provide both insights into and potential indicators of important changes in information-seeking patterns during pandemics such as COVID-19. We asked four questions: (1) How has information seeking changed over time? (2) How does information seeking vary between regions and states? (3) Do states have particular and distinct patterns in information seeking? (4) Do search data correlate with—or precede—real-life events?

Methods: We analyzed searches on 38 terms related to COVID-19, falling into six themes: social and travel; care seeking; government programs; health programs; news and influence; and outlook and concerns. We generated data sets at the national level (covering January 1, 2016, to April 15, 2020) and state level (covering January 1 to April 15, 2020). Methods used include trend analysis of US search data; geographic analyses of the differences in search popularity across US states from March 1 to April 15, 2020; and principal component analysis to extract search patterns across states.

Results: The data showed high demand for information, corresponding with increasing searches for *coronavirus* linked to news sources regardless of the ideological leaning of the news source. Changes in information seeking often occurred well in advance of action by the federal government. The popularity of searches for unemployment claims predicted the actual spike in weekly claims. The increase in searches for information on COVID-19 care was paralleled by a decrease in searches related to other health behaviors, such as urgent care, doctor's appointments, health insurance, Medicare, and Medicaid. Finally, concerns varied across the country; some search terms were more popular in some regions than in others.

Conclusions: COVID-19 is unlikely to be the last pandemic faced by the United States. Our research holds important lessons for both state and federal governments in a fast-evolving situation that requires a finger on the pulse of public sentiment. We suggest strategic shifts for policy makers to improve the precision and effectiveness of non-pharmaceutical interventions and recommend the development of a real-time dashboard as a decision-making tool.

(J Med Internet Res 2021;23(5):e22933) doi:10.2196/22933

XSL•FO
RenderX

KEYWORDS

Google Trends; coronavirus; COVID-19; principal component analysis; information-seeking trends; information retrieval; trend; infodemiology; infoveillance; virus; public health; information seeking; online health information

Introduction

Problem Statement

Over a period of a few months, the COVID-19 pandemic dramatically and rapidly changed the lives of most Americans. With the spread of SARS-CoV-2, people faced extreme uncertainty due to evolving information, calls for behavior change, and economic shocks. As US states move through different stages of the response and plan a path back to normalcy, it is important to develop ways to gather and measure insights into how Americans are responding to the challenges and uncertainties posed by the pandemic [1]. These insights will help policy makers know what to focus on as the United States continues to navigate the COVID-19 outbreak. In particular, it is important to understand the changes and patterns in the information seeking of Americans over the course of the pandemic, both as a window into key concerns as they emerge and as potential signals of actual behavior change.

Collecting data on population responses to a crisis poses a host of methodological challenges [2,3]. Surveys are a commonly used way to gather information on how people are reacting to a health crisis [4]. Several surveys have been deployed to measure Americans' reactions to COVID-19, and they offer insights into the concerns and behaviors of many Americans [5,6]. However, traditional large-scale surveys are expensive and time-consuming; also, due to the rapidly changing nature of the pandemic, it is difficult to capture time-sensitive relevant data [4]. Further, surveys require a baseline understanding of the context of the participants before the events of interest. However, recent research [7] shows that many surveys administered during the COVID-19 pandemic either do not have a baseline or have an inaccurate baseline that relies on the participants' limited ability to recall the answers to retrospective questions. Another option is to use health information systems, generating insights from existing large-scale health data such as electronic patient records and surveillance systems [8]. However, working with this type of data also has limitations: it requires multiple data sets that cannot easily be merged, it relies at times on outdated data, and it requires complex methodological approaches to analysis [9].

An alternative approach to data collection on pandemic response is to examine web-based information-seeking behavior through Google Trends, a source of data on trends in web-based information-seeking [10].

Background

Google Trends and Health Phenomena

Google Trends data have been leveraged extensively in health behavior research by using trends in web-based search queries as proxies for changes in human behavior [11]. Released in 2006, these data became more influential when Ginsberg et al [12] showed a predictive approach that could "make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users." Since then, there have been suggestions to improve the predictive approaches after an overestimation of influenza metrics during the 2012-2013 season [13]. Outside of influenza predictions, Google Trends data have also been used to analyze a wide array of health behavior topics, such as the relationship between media coverage and information seeking around major infection events, the impact of local abortion policies on information seeking about abortion, the influence of influenza concerns on travel, and temporal patterns in interests in healthy behaviors such as dieting [14-17]. In 2014, Nuti et al [3] found that 60% of Google Trends research was focused on infectious diseases and general population behavior, which suggests that it may be a good tool to investigate COVID-19 trends.

Google Trends data are also useful for studying health information- and care-seeking patterns. Initial research using Google Trends data in the United States showed declining interest in imaging diagnostics [18] and surgeries [19] alongside a steady rise in interest in telehealth and telemedicine [20]. Other studies focused on COVID-19 and Google Trends data showed increases in general virus information seeking; however, to date, most studies have explored a singular issue over time or in a specific geographic region, without examining comprehensive search pattern correlations or geographical heterogeneity and their potential implications [21-23]. Recent research also shows "statistically significant correlations between Google Trends and COVID-19 data" [24]. Thus, Google Trends offers immense potential for generating temporally and geographically specific insights into the beliefs, behaviors, and actions of various communities in their response to the COVID-19 pandemic.

Health Information Seeking

Health information seeking is an important step in the care seeking journey, given that "one in three US adults use the internet to diagnose or learn about a health concern" [25]. Health information seeking has been extensively studied, with several studies identifying antecedents and reasons for individual health information-seeking behaviors. Factors that influence health information seeking include the subjective norms of a society or group of which the individual is a member [26], emotional responses and social context [27], and media-driven triggers [28]. Research on web-based health information seeking suggest certain reasons why individuals seek information on the web, including preparing for a doctor's appointment [29] and anxiety about their health [30]. These studies are useful for some applications, such as improving the quality of information campaigns; however, they are general frameworks and are not disease specific.

In this work, we address information seeking from an infodemiology perspective, following seminal work by Eysenbach [30,31] that defines infodemiology as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the

XSL•FO
RenderX

ultimate aim to inform public health and public policy". We focus on the topics people search for (the demand side of infodemiology, in this case represented by Google Trends search queries) as opposed to what people write on the web (the supply side). Some infodemiology demand research focuses only on specific groups of people, such as the information-seeking behavior of clinical providers [32]. Our work instead examines the information seeking of the general population as measured by web-based search activity to capture wider public health implications. We also include a diverse set of socioeconomically related queries to address the multidimensional nature of the effect of COVID-19 and its intersection with nonpharmaceutical interventions. Analytically, we address three of four categories of analysis identified in a systematic review of Google Trends infodemiology by Mavragani et al [33]; we (1) measure the general web-based interest of several COVID-19– related themes, (2) detect variations and seasonality across geographic regions with comparisons to actual events, and (3) correlate search queries among them to understand the underlying patterns. The fourth category involves actual prediction and forecasting, which we do not address in this paper. In addition to specific public health and policy proposals, we emphasize the potential of an infoveillance [34] approach—using web-based information for public health surveillance—and propose a dashboard informed by this work.

Research Questions

We created a curated list of search queries and used Google Trends to analyze changes in information seeking related to the COVID-19 pandemic over time and across geographic areas in the United States. We answered four broad questions regarding information seeking around COVID-19:

First, how is information seeking changing over time? Specifically, what is the relationship between the COVID-19 outbreak and internet searches related to health care seeking, government support programs, media sources of different ideologies, planning around social activities, travel, and food, and new COVID-19–specific behaviors and concerns? We would expect certain types of searches (eg, health care–related inquiries, government safety net programs, web-based food delivery) to become more popular, while others would be expected to fall in popularity (eg, nearby bars, travel plans). These trends provide clear contextual information on the changing interests and concerns of Americans as the pandemic progresses, and we can also compare the timing of these changes with real-world actions and events such as policy announcements.

Second, what geographic variation exists in information-seeking behavior? Specifically, how does the popularity of search terms differ across states and regions? Given the immense heterogeneity across the United States, both culturally and in terms of COVID-19 caseload and response, we would expect states to differ in the most-sought information. Observed geographic differences in information seeking add granularity to the information-seeking context and provide an opportunity to develop hypotheses around why these differences exist.

Third, do states have particular and distinct patterns in information seeking? Specifically, which searches correlate with each other at the state level? Using machine learning methods, specifically unsupervised learning, we can construct different typologies of geographic areas so that those whose search trends provoke concern for public health can be better targeted with information and other interventions.

Fourth, do Google Trends data correlate with and potentially precede real-life events? If Google Trends data can be "predictive" of real-life events (eg, unemployment rates), it provides further validation for this method as a window into the behaviors of Americans during the pandemic.

Methods

Google Trends is a free web-based source of data on trends in web-based information seeking across geographies (country, region, city, and designated market area/metropolitan area where this exists) and over time (since 2004). In 2019, a comprehensive methodological framework was published to standardize approaches to Google Trends research [35]. This work follows that framework, which provides specific criteria for selecting the keywords, geographical regions, and time period for analysis. We also combine several keywords that represent a similar topic into a single query.

Keywords and Search Queries

Google Trends accepts a single word or a phrase, and several of these can be combined into a single Trends query by joining with a "+", which functions like an "OR." This study used 38 queries of up to 5 words/phrases each (Table 1). We generated and curated search terms and evaluated the popularity/relevance and data quality of each query in representing different aspects of the response to and impact of COVID-19–related search activity on Google Trends. The list of search terms was derived from the most commonly searched terms associated with COVID-19 on Google Trends [10] and was refined to be more specific as recommended by Tran et al [36].

Each query was then categorized into one of six emergent themes: care seeking, government programs, health programs, news and influence, outlook and concerns, and social and travel (Table 1). These themes were developed to comprehensively capture the effects of and responses to the pandemic: care seeking and health programs capture how people look for care; news and influence captures the sources of information; and outlook and concerns and social and travel capture the economic and nonpharmaceutical preventive intentions. For the news and influence theme, we selected media outlets that are representative of both sides of the ideological spectrum based on a Pew Research survey conducted in 2014 that measured the ideological leanings (left or right) of the audiences of different news media sources [37]. All search data were extracted on April 15, 2020.

XSL•FO
RenderX

Table 1. Final list of search queries grouped into emergent themes (38 queries with 1-5 combined terms).

Category/ theme	Queries
Care seeking	• *coronavirus symptoms* • *coronavirus testing near me + coronavirus testing center near me + coronavirus test* • *doctor appointment* • *coronavirus afford doctor + coronavirus uninsured + coronavirus medical bill* • *Coronavirus can i see a doctor + coronavirus can i get a test + coronavirus are tests available* • *doctor open + doctor office open* • *urgent care near me*
Government programs	• *disability benefits + apply benefits + food stamps + wic* • *government aid* • *recession + stock market crash + economic downfall + bear market* • *small business loans* • *stimulus check* • *paycheck protection program*
Health programs	• *health insurance* • *health insurance + medicare + medicaid* • *medicaid* • *medicare*
News and influence	• *chinese virus* • *coronavirus cnn + coronavirus msnbc + coronavirus nbc news + coronavirus cbs news* • *coronavirus hoax + coronavirus fake news* • *coronavirus infowars + coronavirus breitbart + coronavirus glenn beck + coronavirus the blaze* • *coronavirus washington post + coronavirus new york times + coronavirus npr* • *coronavirus fox news + coronavirus drudge report*
Outlook and concerns	• *can't pay rent + how pay rent + behind on rent + can't pay mortgage* • *hoarding + hoard* • *how can i stop coronavirus* • *how to make coronavirus mask* • *how to stockpile + buy in bulk + bulk order* • *sick days + sharing sick days + no sick days + sick leave + paid time off* • *social distancing* • *sold out + stock out + stockout + stockpil* • *unemployment benefits + unemployment application + file unemployment + apply unemployment + layoffs*
Social and travel	• *bar closed + restaurant closed* • *bar near me + restaurant reservation + local happy hour + ladies' night* • *bar + restaurant + happy hour + pub + house party + party ideas* • *cheap flights + travel destinations + flight deals + vacation deals* • *food delivery + grocery deliveries + takeout + curbside + online food order* • *house party + party ideas*

Data Collection

Data from Google Trends [10] were extracted through an open-source third party Python application programming interface, PyTrends. Google provides the relative search values (RSVs) for each query on a scale from 0 to 100, representing a normalized value.

Three primary data sets were developed from Google Trends: one at the national level and the other two at state level, capturing all 50 states and the District of Columbia.

The national data set contained weekly RSVs for the United States for each of 38 queries for the 224 weeks between January 1, 2016, and April 15, 2020. This enabled us to conduct trend analysis for each query independently, tracking its relative popularity in the entire United States compared with its most popular week over these 224 weeks. For each query, the week when the query was most popular (as a percentage of all queries in that week) was scored as 100 by Google Trends, and all other weeks were scored relative to this week.

One of the state-level data sets contained weekly RSVs for each query for each state for the 16 weeks between January 1 and April 15, 2020, while the other contained aggregated RSVs for each query for each state over the whole period of March 1 to April 15, 2020. For each of these data sets, RSVs for a specific query were always expressed as search popularity relative to other weeks, other states, or both, but never relative to other search queries (ie, our approach factored out absolute differences in popularity between queries).

Finally, two additional, secondary data sets were referenced, comparing Google Trends against new monthly Medicaid applications and weekly initial unemployment claims, from the public Medicaid web-based database (June 2017 to January

XSL•FO
RenderX

2020) [38] and US Department of Labor data (January 2016 to April 2020) [39], respectively.

Analysis

Geographically, states were grouped by federal region and division based on US census data [40]. We used Python for data processing and Tableau for exploratory data analysis and visualization [41]. We used the packages of Python and R (R Foundation for Statistical Computing) to perform pairwise correlations (Pearson, pairwise complete) and unsupervised learning by principal component analysis (PCA), respectively [42]. We conducted pairwise correlations to visualize and quantify associations between the individual queries at the state level; we then applied dimensionality reduction through PCA to reveal groups of queries that are often searched together and to identify top states with search patterns of interest, either supporting or undermining the fight against the COVID-19 epidemic. Missing values were removed for correlation and PCA analysis (ie, we retained all 51 geographies and removed any queries that had missing values for any states). Nine queries had missing values: *coronavirus infowars + coronavirus breitbart + coronavirus glenn beck + coronavirus the blaze, how can I stop coronavirus, coronavirus can I see a doctor + coronavirus can I get a test + coronavirus are tests available, coronavirus afford doctor + coronavirus uninsured + coronavirus medical bill, bar closed + restaurant closed, government aid, doctor appointment, doctor open + doctor office open*, and *can't pay rent + how pay rent + behind on rent + can't pay mortgage*. This reduced the number of queries for analysis from 38 to 29.

For pairwise correlations, we filtered for any correlations of low to very high (positive or negative) correlation (ie, >0.3 or <–0.3) according to the guideline proposed by Mukaka [43]. We then focused only on the most prominent (moderate and high) scales. For PCA, only loadings with loading scores higher than 0.2 or lower than –0.2 were used to explain each component (PCA loadings take values from 0 to 1).

To calculate the increase in search popularity between January and March 2020, we took the mean RSV for March 2020 and divided this by the mean RSV for January 2020 and expressed this ratio as a percentage change. We chose January and March to capture the largest shifts in RSV based on the progression of the SARS-CoV-2 virus. Data for the extra 2 weeks in April were used to monitor which trends persisted or dissipated after large shifts in March.

Finally, to compare Google Trends to real-life phenomena, we visualized data on actual weekly initial unemployment claims and new applications for Medicaid, then quantified the correlations between each phenomenon and the corresponding Google Trends query.

Results

Magnitude of Shifts in Information Seeking

National trends show shifts in information seeking related to health and lifestyle activities.

Analysis of all queries showed substantial shifts in RSVs across all thematic categories (Figure 1) predominantly in March 2020, with trends persisting or stabilizing in early April.

For the care seeking theme, the RSVs of COVID-19–related queries increased substantially (*coronavirus symptoms* and *testing centers*) while the RSVs of general health-seeking queries (*urgent care* and *doctor appointments*) declined in late March/early April, suggesting a nuanced story of care seeking (Figure 1). This trend was matched by a decline in the RSVs for all queries in the health programs theme (*health insurance*, *medicaid*, and *medicare*) by 18%, 23%, and 26%, respectively. Comparison with the same time window in preceding years confirmed that this finding was an anomaly against historical seasonality, where health program searches peak and drop in November and December, not in March and April.

News and influence searches related to COVID-19 saw notable hikes in RSV for both left-leaning and right-leaning media. RSV for far-right/alt-right media outlets (see the *Methods* section for definitions based on the Pew Research survey) also increased. Simultaneously, RSV for *coronavirus fake news* and *coronavirus hoaxes* surged 38-fold (Figure 1).

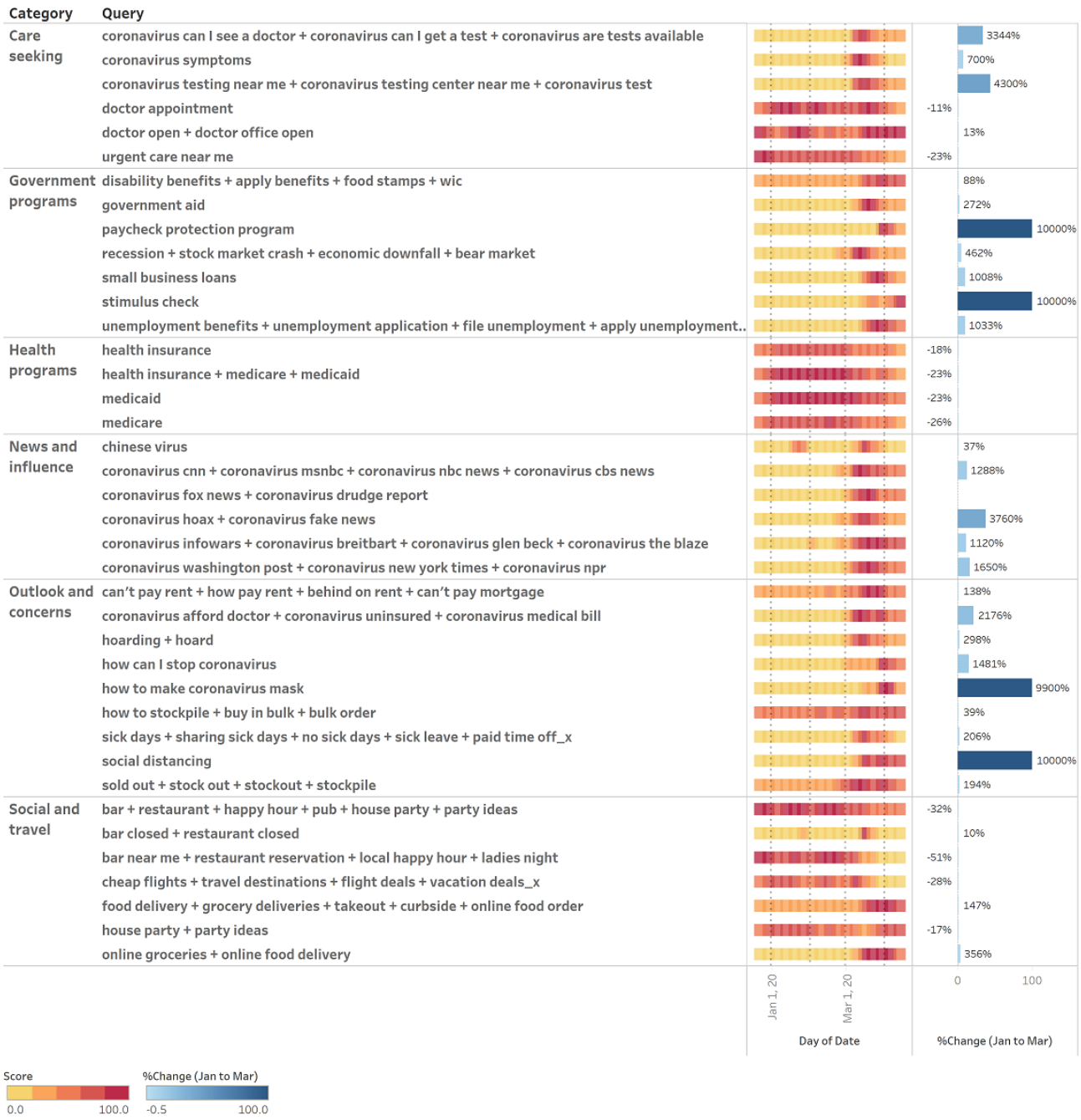
For the outlook and concerns theme, RSVs for new behavioral concepts gained immense popularity: *social distancing* and *how to make masks* spiked by 100-fold and peaked in March, as did time-sensitive concerns such as *hoarding* and *can't pay rent*. In contrast, searches for *stock outs/sold out* and *coronavirus medical bill/affordability* remained at high levels into early April, hinting at potential differences in long-term versus short-term concerns (Figure 1).

Within the social and travel theme, search trends were aligned with the new norm of social distancing and also signaled potential drops in business for the travel and service industries. The RSVs for *online groceries + food delivery* and *food delivery + grocery deliveries + takeout + curbside + online food order* tripled and doubled, respectively, while the RSVs for *party ideas*, *nearby bars and restaurants*, and *cheap flights/travel* all dropped by between 17% and 51%. This shift is noteworthy for *nearby bars and restaurants*, as it is a reversal of an upward trajectory that had recently peaked in December 2019 (Figure 1).

For the government programs theme, its relative popularity increased for new COVID-19–specific packages (*stimulus check* and *small business loans*) as well for unemployment benefits, which is a more mature program. Notably, the RSVs for *disability/food stamps* and *government aid* had the lowest increases in RSVs among government programs. However, relative interest in programs focused on individuals—*stimulus check* and *unemployment application*—maintained momentum through April 15 (Figure 1).

XSL•FO
RenderX

Figure 1. National search pattern for each query in every theme by relative search value for 16 weeks between January 1 and April 15, 2020. The right panel shows the change in monthly average relative search value between January and March, capped at 10,000% for queries with extremely high changes.



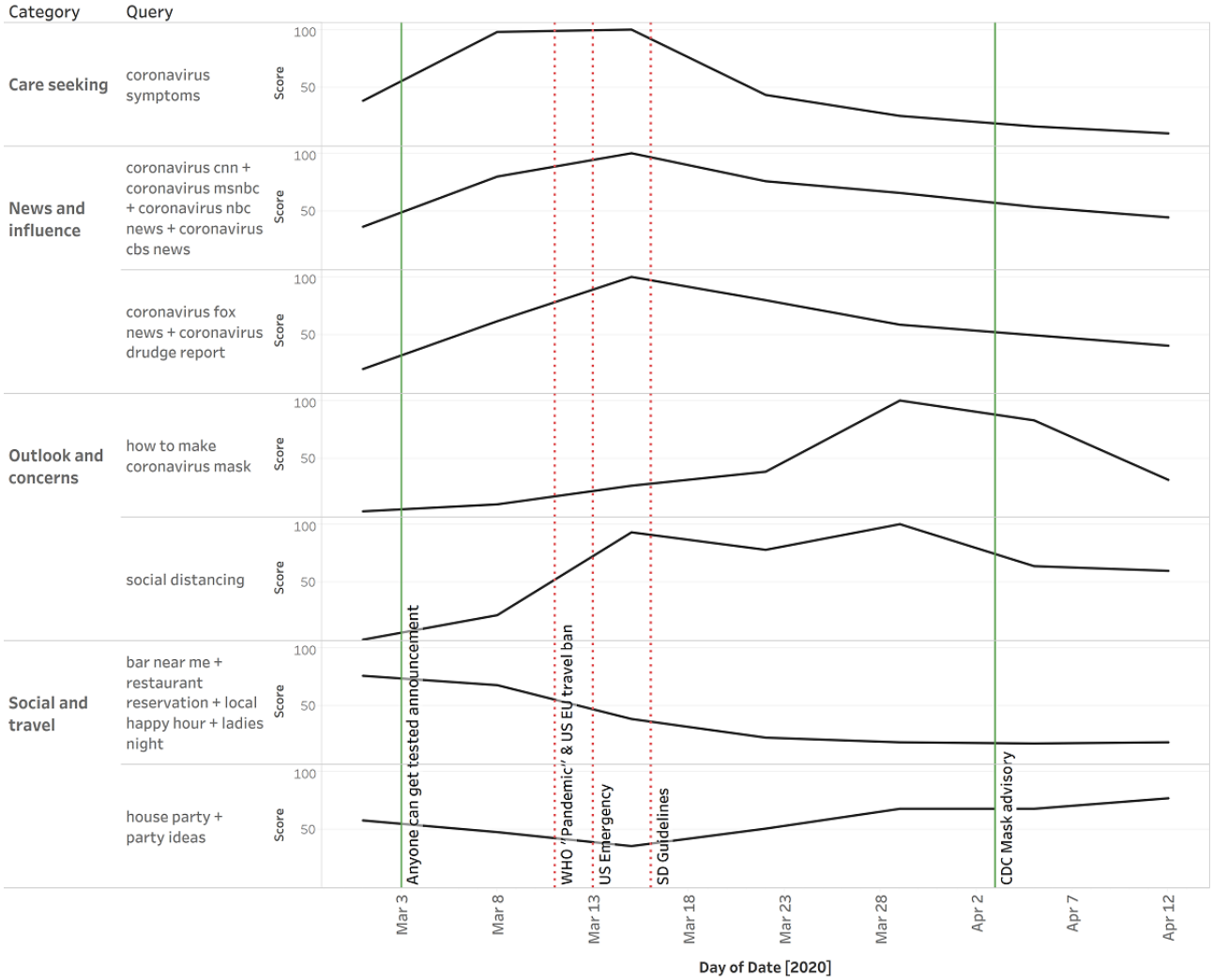
Timing of Shifts in Information Seeking

Shifts in information seeking occurred earlier than federal government policy action. We would have expected the directives from the federal government, which formulates national policies, to be issued early and to precede changes in search patterns and behaviors. However, the Google Trends data show that the relative popularity of queries related to non-pharmaceutical interventions (NPIs) was already shifting substantially, and in some cases peaking, several days ahead of major federal government policies and NPI action (Figure 2). For example, RSVs for *social distancing* and *bar/restaurant nearby* were quickly increasing and decreasing, respectively, by March 8—3 days before the World Health Organization declared COVID-19 a pandemic, 5 days before the US federal government's national emergency declaration, and 8 days before the government released official social distancing guidelines. Furthermore, *how to make a coronavirus mask* was already relatively popular (RSV >50) by March 23, exactly 12 days before the US Centers for Disease Control and Prevention issued an advisory promoting the use of masks for the general public.

On the other hand, the RSV trends for news on COVID-19 and COVID-19 symptoms were in sync, matching each other regardless of the ideological leaning of the news source. This suggests that searches for disease characteristics were highly correlated with searches for news coverage of the disease.

XSL•FO
RenderX

Figure 2. National trends for Google Trends nonpharmaceutical intervention–related queries compared to actual government and public health nonpharmaceutical interventions at the US federal level between March 1 and April 15, 2020. CDC: US Centers for Disease Control and Prevention; EU: European Union; SD: social distancing; WHO: World Health Organization.



Correlation Between Information Seeking and Unemployment and Medicaid Claims

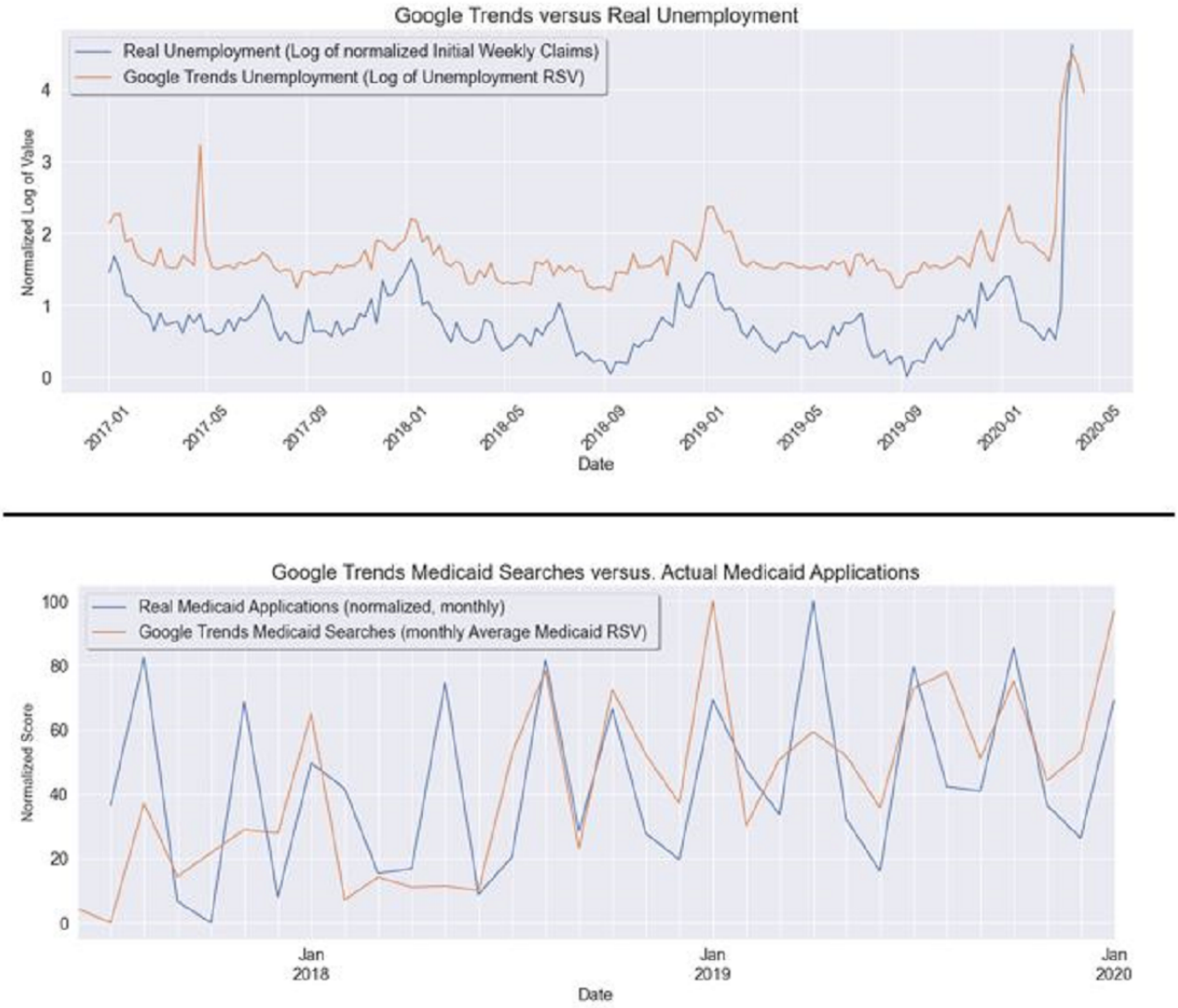
Google Trends were correlated with unemployment claims and Medicaid applications. We found high correlations between specific Google Trends queries and corresponding phenomena in the real world both before the epidemic (for unemployment and Medicaid) and during the epidemic (for unemployment). Two examples were selected: initial weekly unemployment claims and new monthly Medicaid applications. Figure 3 shows the comparative results with corresponding Google Trends queries over time.

There was a very high positive correlation of 0.96 between the previous week's Google Trends RSV for unemployment applications and the actual weekly initial unemployment claims normalized to 0-100 (seasonally adjusted as recommended by the Bureau of Labor Statistics to eliminate seasonal spikes and enable easy detection of nonseasonal anomalies such as COVID-19–related spikes). For Medicaid, there was a moderate positive correlation of 0.55 between the average monthly Google Trends RSV for Medicaid, lagged one month, and the number of new applications for Medicaid each month.

However, we noted that this approach was only modestly successful for more complex cases in which the link between search popularity and actual behavior is more difficult to imagine. A notable example is stock market prices (Dow Jones Industrial Average) and the Google Trends query for *recession/stock market crash*, which showed no correlation.

XSL•FO
RenderX

Figure 3. National-level multivariate trends for new weekly unemployment (January 2016 to April 2020) [39] and new monthly Medicaid applications (July 2017 to January 2020) [38]. Top: Google searches for unemployment applications and actual initial weekly unemployment claims normalized from 0-100 over time (with seasonal adjustment). Bottom: Monthly average Google searches for Medicaid and actual new applications for Medicaid normalized from 0-100 (lagged by 1 month). RSV: relative search value.



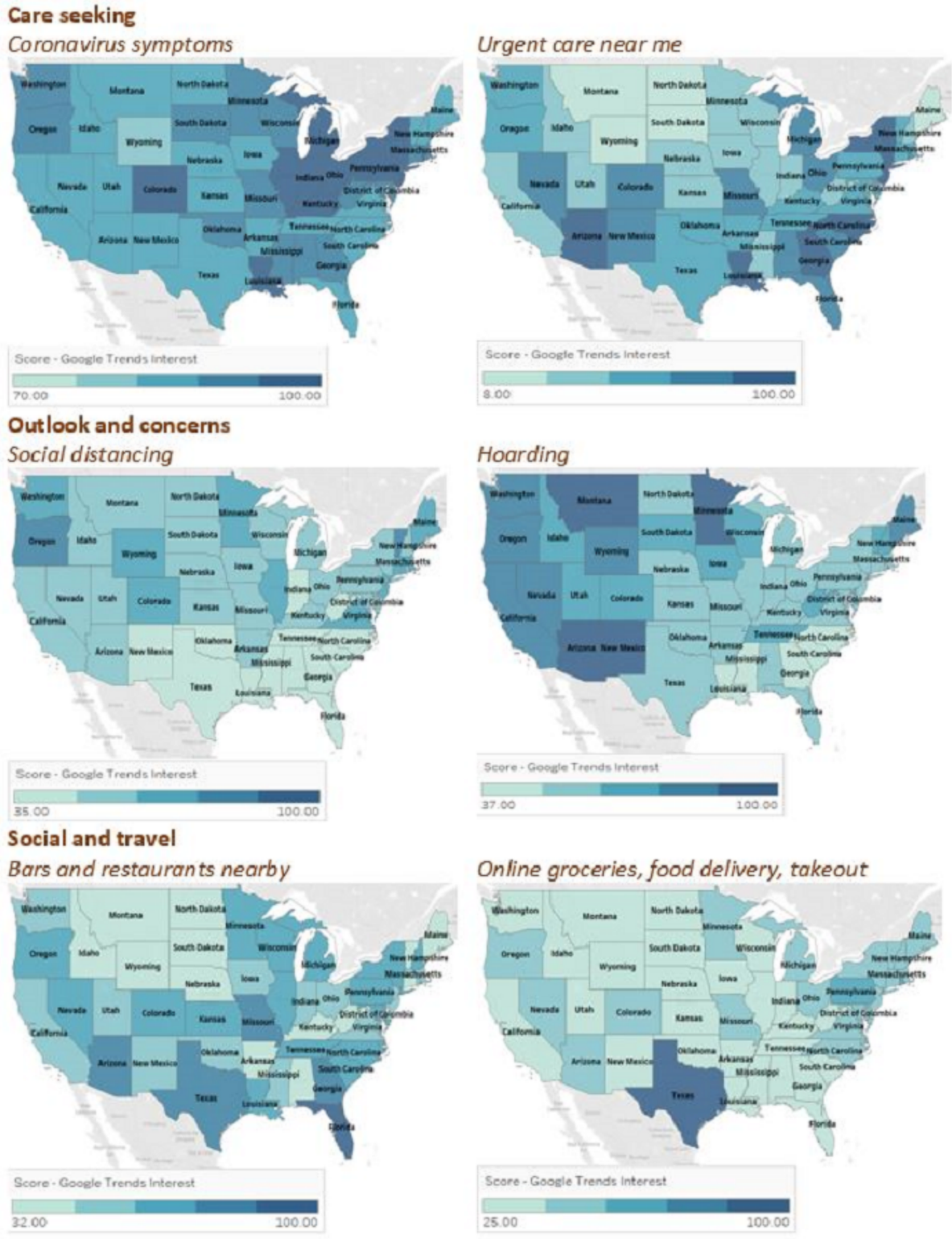
Differences in Search Patterns Across States and Regions

Differences in search patterns suggest varied responses to COVID-19 across states and regions. Nationally, the largest jumps in RSV were concentrated between March 1 and April 15; therefore, we focused on this window to investigate state-by-state differences in information seeking. Here, we compared the relative popularity of a search query across states in a single time period rather than across time.

The findings showed similar levels of popularity of searches for *urgent care near me* in the South and Northeast but a difference in the relative popularity of searches for health programs (*health insurance*, *Medicare*, and *Medicaid*). Care seeking searches were most popular in the South (Louisiana, Georgia, and North Carolina); the Northeast (New York and New Jersey); and in Indiana, Illinois, and Arizona during this specific window (Figure 4). Although New York and New Jersey were already COVID-19 hotpots, this period also coincided with sharp increases in the number of cases for Louisiana, Georgia, and North Carolina.

XSL•FO
RenderX

Figure 4. Geographical variations across states from March 1 to April 15, 2020, for the care seeking, outlook and concerns, and social and travel themes. The figure shows the two queries that were most representative of each category.

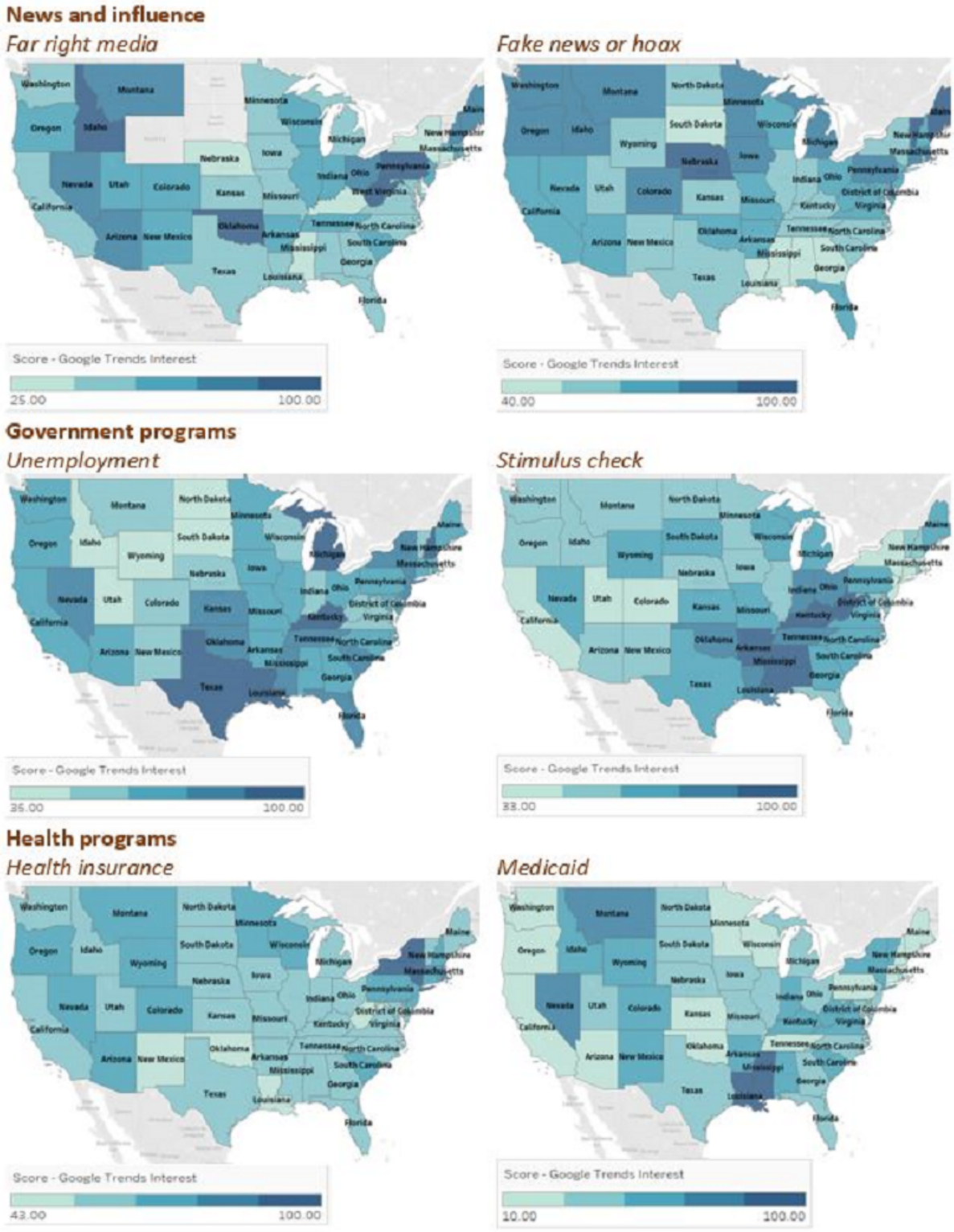


RSVs for *social distancing* were generally high in the Northeast and West but generally low in the South, while searches for *hoarding* were relatively more popular in Alaska, New Mexico, Minnesota, Arizona, and Montana than in other states during this period.

Regarding the news and influence theme, searches for *far-right/alt-right* and *coronavirus* were most popular in West Virginia, Oklahoma, Idaho, and Pennsylvania, while *fake news coronavirus* searches were most popular in DC, Vermont, Alaska, Maine, and Nebraska (Figure 5). Additionally, *stimulus check* had particularly high RSVs in Southern states (Figure 5). Although *health insurance* had a high RSV in the Northeast (New York, Massachusetts, and Vermont), *Medicaid* had higher RSVs in Southern states, including Louisiana and Mississippi (Figure 5).

XSL•FO
RenderX

Figure 5. Geographical variations across states from March 1 to April 15, 2020, for the news and influence, government programs, and health programs themes. The figure shows the two queries that were most representative of each category.



Summarizing Differences in Search Patterns Using Correlation and PCA

Much variability in information seeking between states can be explained by a few components.

Although the thematic categories in the above analysis give insights on the relative popularity of queries over time and across geographic regions, we also conducted pairwise correlation and PCA analyses to understand how the queries were correlated. First, correlation analysis showed that states with high RSVs for *coronavirus symptoms* also tended to have high RSVs for *urgent care* and *test centers*, suggesting a relationship between awareness of the disease and potential intention to seek care. Second, the correlation showed that states with high RSVs for *social distancing* also tended to have high RSVs for *coronavirus* and all types of news media regardless

XSL•FO
RenderX

of ideological leaning, *recession/stock market crash*, and *sick days/leave*, while tending to have low RSVs for safety net programs such as *medicare* and *stimulus check*. This finding suggests that news sources and economic factors play a role in the levels of interest in, awareness of, and potential adoption of social distancing behavior.

The PCA results are shown in Figure 6, with all 38 queries summarized as the top two components that explain 43% of the variation (differences or patterns) in the data—25% from Component 1 and 18% from Component 2. The two components were assigned labels based on the search patterns they showed.

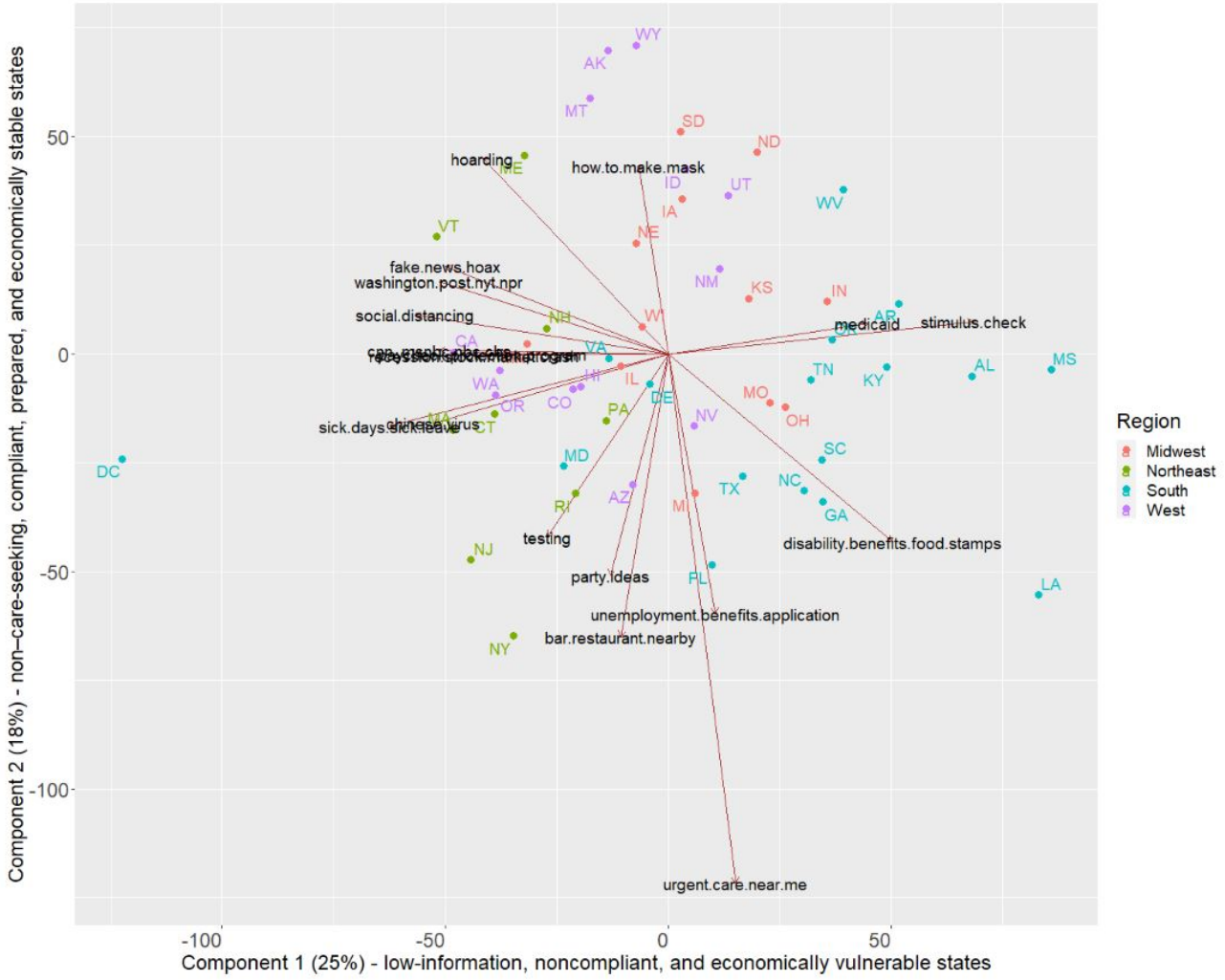
PCA revealed search patterns related to economic vulnerability and searching for information from news and media sources. It also highlighted how state search patterns were related to other concepts, such as compliance with social distancing or related policies such as mask wearing; preparation for emergencies (hoarding); and care-seeking concepts, such as searching for urgent care.

Specifically, Component 1 explained 25% of the variation in the data (Figure 6). This component represented potentially low-information, noncompliant, and economically vulnerable states. These terms are defined as follows: *low-information*: low association with searches for any news source, whether real or fake; *noncompliant*: low association with searches for *social distancing*; and *economically vulnerable*: high association with searches for *disability/food stamps* and *stimulus check*. The states with the highest scores (ie, those that exhibited this search pattern most strongly) were Mississippi, Louisiana, Alabama, Arkansas, and Kentucky. Additional clusters of states within Component 1 also had high scores for *urgent care nearby* and *unemployment application* (Florida and Michigan) as well as *disability/food stamps* (Georgia, North Carolina, and South Carolina).

Component 2 explained 18% of the variation in the data (Figure 6). This represented potentially non–care-seeking, compliant, prepared, and economically stable states. These terms are defined as follows: *non–care-seeking*: low association with searches for *urgent care nearby*; *compliant*: low association with searches of *nearby bars/restaurants*; *prepared*: high association with searches for *how to make a mask* and *hoarding*; and *economically stable*: low association with searches for *unemployment* and *disability/food stamps*. The states exhibiting this trend were Wyoming, Alaska, Montana, North Dakota, and South Dakota.

XSL•FO
RenderX

Figure 6. Scatterplot of state principal component analysis loadings and scores for the first two components, with the top queries shown as arrow vectors. Each arrow represents the relative weight of each query, and the direction indicates the points to the states that most exhibit this search pattern. The arrow direction measures correlation; arrows in the same direction are highly positively correlated, while divergent arrows in opposite directions are highly negatively correlated. Component 1 (x-axis) and Component 2 (y-axis) explain 25% and 18% of the variation (differences or patterns) in the data, respectively.



Discussion

Principal Findings

We studied the demand side of the infodemiology [31] of COVID-19 using a curated set of queries from Google Trends grouped into emergent themes. Our findings show substantial changes in COVID-19 information seeking at the national level, particularly in March 2020, suggesting a hyperawareness and desire for information about both COVID-19 and its corresponding novel behavioral concepts, namely social distancing and mask wearing. This is in line with other studies showing spikes in information seeking during disease outbreaks [44]. The trends also mirrored the rapid changes in the way people eat, travel, and socialize. The high demand for information corresponded with increasing searches for news sources and coronavirus, regardless of ideological leaning, including searches for *coronavirus fake news* + *coronavirus hoax*. It is not conclusive whether this indicates curiosity or an earnest belief in the existence of "fake news." Still, this finding underscores the critical and timely role that news sources play in providing information during a pandemic and why this information must be correct and trustworthy. This complements other Google Trends studies [45,46] that showed that information seeking and collective attention toward COVID-19 can, at times, be driven by media coverage more than epidemic trends.

Effective communication during health crises is critical and, if not done well, leads to "public confusion and misunderstanding" with negative public health consequences, as happened with COVID-19 in the United States [47,48]. Our study found that changes in information seeking often occurred well in advance of action by the federal government. Tracking search patterns for social distancing or mask wearing could show public health authorities what interventions people are aware of and might accept, as well as the best time to start talking about them, especially in rapidly evolving situations. Previous studies showed that attention paid to COVID-19–related information spiked sharply but quickly saturated despite high media coverage [44] as the pandemic continued to spread. Therefore, the right interventions, if poorly timed, could be too early to be acceptable, or, as occurred in the case of the COVID-19 pandemic, too late to be optimally effective.

XSL•FO
RenderX

The decrease in popularity of searches related to urgent care, doctors' appointments, and health insurance/Medicare/Medicaid is consistent with the decline in health seeking that occurs during pandemics: prospective patients for other diseases have higher risk perception of hospital-based transmission of COVID-19, which reduces their health-seeking behavior [49]. This was especially true in the United States, where hospitals prioritized COVID-19 and relegated other services [50]. This finding can be an early-warning indicator of reduced health-seeking behavior, especially for vulnerable communities with underlying health risks. It can be used to prioritize and target these communities with messages about the continued importance of health care for conditions and symptoms not related to COVID-19.

State-level differences in search patterns confirmed the trajectory of the pandemic and informed our potential hypotheses about the regional and structural drivers of these differences. For example, in New York and New Jersey, two states where the pandemic quickly accelerated and where, coincidentally, health insurance marketplaces are state-run rather than federally run, people were performing searches related to urgent care and health insurance [51]. Residents of southern states tended to search for several potentially worrying factors in the fight against COVID-19, including searches related to urgent care and Medicaid and searches for stimulus checks, indicating the need for a financial safety net. The search for *hoarding* was popular in states with either large land areas (and less dense populations) or especially extreme weather conditions—Alaska, New Mexico, Minnesota, Arizona, and Montana. This may reflect heightened expectations of scarcity of the populations of these states during a pandemic, as these factors likely make these states hard to reach in the case of disrupted supply chains. These hypotheses need to be validated; however, they provide a starting point for anticipating the trajectory of the next pandemic or national emergency.

Using correlation and PCA, we identified economic and financial factors, access to information, and interest in social distancing as key variables in describing states' information-seeking patterns on COVID-19. Using this premise, Mississippi, Louisiana, Alabama, Arkansas, and Kentucky were identified as low-information, noncompliant, and economically vulnerable states during the time window of the analysis (March and April 2020). Thus, the states that are most vulnerable economically or in terms of the social safety net are also the least informed and show the least search-related interest in social distancing, a key intervention to prevent COVID-19. This insight can be used to increase awareness of social distancing and its benefits while targeting and prioritizing resources to support these states, including increased testing, health system capacity support, and economic relief measures.

To capture the evolving information and insights available through Google Trends, we propose a real-time dashboard to track trends, geographical variations, and patterns of interest regarding the epidemic. This proposal is in line with evidence that Google Trends data are statistically significantly correlated with COVID-19 data [24]. The queries used in this study can be used as a starting point or baseline, with additional features added if needed. As the epidemic continues and potentially gives way to a second wave, this dashboard will follow how specific queries change over time and which states demonstrate concerning search patterns using the pairwise correlation and PCA approaches. Depending on the search patterns identified, policy makers can then design or improve interventions as well as allocate resources to target states.

Limitations

We made some assumptions and choices that may result in limitations depending on the use case. First, the selection of keywords and queries was an iterative process because it is difficult to know a priori what search terms best capture a desired concept. Analytically, we used weekly RSVs as opposed to daily RSVs, resulting in some loss of granularity. We also used the state as our main unit of analysis and not the city or designated market area level because the COVID-19 response was organized at the state level and lower levels had more missing values. However, at state level, there were a few missing values that were included in the trend and geographical analyses but removed for the pairwise correlation and PCA analyses.

While valuable insights can be derived, a major limitation of Google Trends is that it will always be a measure of search patterns and not the actual corresponding behaviors; therefore, inference and prediction must always come with this caveat. Finally, this analysis is limited by specific assumptions of Google Trends, which include pulling the data from only a sample and not the whole database of searches and providing it in the form of RSV instead of absolute search volumes per geographic region. This requires caution when analyzing results from low-volume queries or geographic regions and when making interpretations and conclusions from analyses.

Conclusion

Our work provides insights into, and potential indicators of, important changes in information-seeking patterns during pandemics such as COVID-19 that can be used to inform public health and public policy. High demand for information corresponding with increasing search popularity for COVID-19 from news sources highlights the importance of public health authorities working with media to ensure that information is correct. Decreases in search popularity for health seeking–related searches can be an early warning indicator for policy makers to target areas with serious underlying health risks with messages about the importance of continuing to seek care for other ailments, even during a pandemic. The emergence of economic and financial factors, access to information, or interest in social distancing as important variables suggests potential interventions to increase awareness of social distancing and its benefits, while targeting and prioritizing resources to support states, including increased testing, health system capacity support, and economic relief measures.

We demonstrated that PCA can explain the variation across several queries and multiple geographies. This work underscores the importance of tracking a well-curated set of queries, customized for the topic, and combined to capture all angles of the problem to improve public health in real time.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Survey tool and guidance: rapid, simple, flexible behavioural insights on COVID-19. World Health Organization. 2020 Jul 29. URL: https://apps.who.int/iris/handle/10665/333549 [accessed 2021-04-27]
2. Rubin GJ, Amlôt R, Page L, Wessely S. Methodological challenges in assessing general population reactions in the immediate aftermath of a terrorist attack. Int J Methods Psychiatr Res 2008 Dec 26;17 Suppl 2(S2):S29-S35 [FREE Full text] [doi: 10.1002/mpr.270] [Medline: 19035438]
3. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
4. Rubin G, Bakhshi S, Amlot R, Fear N, Potts H, Michie S. The design of a survey questionnaire to measure perceptions and behaviour during an influenza pandemic: the Flu TElephone Survey Template (FluTEST). Health Services and Delivery Research 2014;2(41). [doi: 10.3310/hsdr02410] [Medline: 25642523]
5. Nelson LM, Simard JF, Oluyomi A, Nava V, Rosas LG, Bondy M, et al. US public concerns about the COVID-19 pandemic from results of a survey given via social media. JAMA Intern Med 2020 Jul 01;180(7):1020-1022 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1369] [Medline: 32259192]
6. Ashley KA, Kearney A, Hamel L, Brodie M. KFF health tracking poll - early April 2020: the impact of coronavirus on life in america. Keiser Family Foundation. 2020 Apr. URL: https://www.kff.org/health-reform/report/kff-health-tracking-poll-early-april-2020/ [accessed 2021-04-27]
7. Sauermann L, Bünning M, Bunnes S, Sauermann A. Problems and pitfalls of retrospective survey questions in COVID-19 studies. Surv Res Methods 2020;14(2):02-14. [doi: 10.18148/srm/2020.v14i2.7741]
8. Salathé M. Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
9. Buckee C. Improving epidemic surveillance and response: big data is dead, long live big data. Lancet Digit Health 2020 May;2(5):e218-e220. [doi: 10.1016/s2589-7500(20)30059-5]
10. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2021-04-27]
11. Kan W, Chien T, Wang H, Chou W. The most cited articles on the topic of health behaviors in Google Trends research: a systematic review. Adv Gen Pract Med 2018 Dec 17:1-7. [doi: 10.25082/agpm.2018.02.001]
12. 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]
13. Martin LJ, Xu B, Yasui Y. Improving Google Flu Trends estimates for the United States through transformation. PLoS One 2014;9(12):e109209 [FREE Full text] [doi: 10.1371/journal.pone.0109209] [Medline: 25551391]
14. Kostkova P, Fowler D, Wiseman S, Weinberg JR. Major infection events over 5 years: how is media coverage influencing online information needs of health care professionals and the public? J Med Internet Res 2013 Jul 15;15(7):e107 [FREE Full text] [doi: 10.2196/jmir.2146] [Medline: 23856364]
15. Reis BY, Brownstein JS. Measuring the impact of health policies using Internet search patterns: the case of abortion. BMC Public Health 2010 Aug 25;10:514 [FREE Full text] [doi: 10.1186/1471-2458-10-514] [Medline: 20738850]
16. Fenichel EP, Kuminoff NV, Chowell G. Skip the trip: air travelers' behavioral responses to pandemic influenza. PLoS One 2013;8(3):e58249 [FREE Full text] [doi: 10.1371/journal.pone.0058249] [Medline: 23526970]
17. Carr LJ, Dunsiger SI. Search query data to monitor interest in behavior change: application for public health. PLoS One 2012;7(10):e48158 [FREE Full text] [doi: 10.1371/journal.pone.0048158] [Medline: 23110198]
18. Adelhoefer S, Henry TS, Blankstein R, Graham G, Blaha MJ, Dzaye O. Declining interest in clinical imaging during the COVID-19 pandemic: An analysis of Google Trends data. Clin Imaging 2021 May;73:20-22 [FREE Full text] [doi: 10.1016/j.clinimag.2020.11.037] [Medline: 33260013]
19. Jella TK, Samuel LT, Acuña AJ, Emara AK, Kamath AF. Rapid decline in online search queries for hip and knee arthroplasties concurrent with the COVID-19 pandemic. J Arthroplasty 2020 Oct;35(10):2813-2819 [FREE Full text] [doi: 10.1016/j.arth.2020.05.051] [Medline: 32534864]
20. Arshad Ali S, Bin Arif T, Maab H, Baloch M, Manazir S, Jawed F, et al. Global interest in telehealth during COVID-19 pandemic: an analysis of Google Trends. Cureus 2020 Sep 16;12(9):e10487 [FREE Full text] [doi: 10.7759/cureus.10487] [Medline: 33083187]
21. Lucas B, Elliot B, Landman T. Online information search during COVID-19. ArXiv. Preprint posted online on April 15, 2020 [FREE Full text]
22. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. Brain Behav Immun 2020 Aug;88:950-951 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.042] [Medline: 32311493]

XSL•FO
RenderX

23. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: evidence from Google Trends. SSRN Journal. Preprint posted online on March 12, 2020. [doi: 10.2139/ssrn.3550008]
24. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 2020 Nov 26;10(1):20693 [FREE Full text] [doi: 10.1038/s41598-020-77275-9] [Medline: 33244028]
25. 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]
26. Chang C, Huang M. Antecedents predicting health information seeking: a systematic review and meta-analysis. Int J Inf Manage 2020 Oct;54:102115 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102115]
27. Pian W, Song S, Zhang Y. Consumer health information needs: a systematic review of measures. Inf Process Manag 2020 Mar;57(2):102077 [FREE Full text] [doi: 10.1016/j.ipm.2019.102077]
28. Hogue MB, Doran E, Henry DA. A prompt to the web: the media and health information seeking behaviour. PLoS One 2012;7(4):e34314 [FREE Full text] [doi: 10.1371/journal.pone.0034314] [Medline: 22509289]
29. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]
30. McMullan R, 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 [FREE Full text] [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
31. 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]
32. Gordon ID, Meindl P, White M, Szigeti K. Information seeking behaviors, attitudes, and choices of academic chemists. Sci Technol Libr 2018 Apr 09;37(2):130-151. [doi: 10.1080/0194262x.2018.1445063]
33. 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]
34. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
35. 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]
36. Tran US, Andel R, Niederkrotenthaler T, Till B, Ajdacic-Gross V, Voracek M. Low validity of Google Trends for behavioral forecasting of national suicide rates. PLoS One 2017;12(8):e0183149 [FREE Full text] [doi: 10.1371/journal.pone.0183149] [Medline: 28813490]
37. Mitchell A, Gottfried J, Kiley J, Matsa KE. Pew Research Center. 2014. URL: https://www.journalism.org/2014/10/21/political-polarization-media-habits/ [accessed 2021-04-27]
38. Medicaid data. Medicaid. URL: https://data.medicaid.gov [accessed 2020-04-15]
39. Unemployment insurance weekly claims data. United States Department of Labor. URL: https://oui.doleta.gov/unemploy/claims.asp [accessed 2021-04-27]
40. Census regions and divisions of the United States. U.S. Census Bureau. URL: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [accessed 2021-04-27]
41. Oliphant TE. Python for Scientific Computing. Comput Sci Eng 2007 May;9(3):10-20. [doi: 10.1109/mcse.2007.58]
42. R Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
43. 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]
44. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
45. Sousa-Pinto B, Anto A, Czarlewski W, Anto J, Fonseca J, 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]
46. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. J Med Internet Res 2005 Jul 01;7(3):e36 [FREE Full text] [doi: 10.2196/jmir.7.3.e36] [Medline: 15998627]
47. Kim DKD, Kreps GL. An analysis of government communication in the United States during the COVID-19 pandemic: recommendations for effective government health risk communication. World Med Health Policy 2020 Aug 27;12(4):398-412 [FREE Full text] [doi: 10.1002/wmh3.363] [Medline: 32904935]

XSL•FO
RenderX

48. Nuzzo JB, Bell JA, Cameron EE. Suboptimal US response to COVID-19 despite robust capabilities and resources. JAMA 2020 Sep 16;324(14):1391. [doi: 10.1001/jama.2020.17395] [Medline: 32936215]
49. Ateev MM, Linetsky D, Hatch H, Cutler D. To the Point (blog). Commonwealth Fund. 2020. URL: https://www.commonwealthfund.org/blog [accessed 2021-04-27]
50. Healthcare facilities: managing operations during the COVID-19 pandemic. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhealthcare-facilities%2Fguidance-hcf.html [accessed 2021-04-27]
51. Lueck S. Adopting a state-based health insurance marketplace poses risks and challenges. Center on Budget and Policy Priorities. 2020 Feb 06. URL: https://www.cbpp.org/research/health/adopting-a-state-based-health-insurance-marketplace-poses-risks-and-challenges [accessed 2020-06-02]

Abbreviations

NPI: nonpharmaceutical intervention
PCA: principal component analysis
RSV: relative search value

Edited by C Basch; submitted 17.08.20; peer-reviewed by L Sbaffi, S Kardes; comments to author 13.10.20; revised version received 18.11.20; accepted 15.04.21; published 03.05.21.

Please cite as:
Mangono T, Smittenaar P, Caplan Y, Huang VS, Sutermaster S, Kemp H, Sgaier SK
Information-Seeking Patterns During the COVID-19 Pandemic Across the United States: Longitudinal Analysis of Google Trends Data
J Med Internet Res 2021;23(5):e22933
URL: https://www.jmir.org/2021/5/e22933
doi:10.2196/22933
PMID:33878015

©Tichakunda Mangono, Peter Smittenaar, Yael Caplan, Vincent S Huang, Staci Sutermaster, Hannah Kemp, Sema K Sgaier. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2021. 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

Mining and Validating Social Media Data for COVID-19–Related Human Behaviors Between January and July 2020: Infodemiology Study

Ashlynn R Daughton[1], MPH, PhD; Courtney D Shelley[1], PhD; Martha Barnard[1], BA; Dax Gerts[1], MS; Chrysm Watson Ross[1,2], MA, MS; Isabel Crooker[1]; Gopal Nadiga[1]; Nilesh Mukundan[1]; Nidia Yadira Vaquera Chavez[1,2], MS; Nidhi Parikh[1], PhD; Travis Pitts[1], MA; Geoffrey Fairchild[1], PhD

[1]Analytics, Intelligence, and Technology, Los Alamos National Laboratory, Los Alamos, NM, United States

[2]Computer Science, University of New Mexico, Albuquerque, NM, United States

Corresponding Author:
Ashlynn R Daughton, MPH, PhD
Analytics, Intelligence, and Technology
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM, 87545
United States
Phone: 1 505 664 0062
Email: adaughton@lanl.gov

Abstract

Background: Health authorities can minimize the impact of an emergent infectious disease outbreak through effective and timely risk communication, which can build trust and adherence to subsequent behavioral messaging. Monitoring the psychological impacts of an outbreak, as well as public adherence to such messaging, is also important for minimizing long-term effects of an outbreak.

Objective: We used social media data from Twitter to identify human behaviors relevant to COVID-19 transmission, as well as the perceived impacts of COVID-19 on individuals, as a first step toward real-time monitoring of public perceptions to inform public health communications.

Methods: We developed a coding schema for 6 categories and 11 subcategories, which included both a wide number of behaviors as well codes focused on the impacts of the pandemic (eg, economic and mental health impacts). We used this to develop training data and develop supervised learning classifiers for classes with sufficient labels. Classifiers that performed adequately were applied to our remaining corpus, and temporal and geospatial trends were assessed. We compared the classified patterns to ground truth mobility data and actual COVID-19 confirmed cases to assess the signal achieved here.

Results: We applied our labeling schema to approximately 7200 tweets. The worst-performing classifiers had F1 scores of only 0.18 to 0.28 when trying to identify tweets about monitoring symptoms and testing. Classifiers about social distancing, however, were much stronger, with F1 scores of 0.64 to 0.66. We applied the social distancing classifiers to over 228 million tweets. We showed temporal patterns consistent with real-world events, and we showed correlations of up to –0.5 between social distancing signals on Twitter and ground truth mobility throughout the United States.

Conclusions: Behaviors discussed on Twitter are exceptionally varied. Twitter can provide useful information for parameterizing models that incorporate human behavior, as well as for informing public health communication strategies by describing awareness of and compliance with suggested behaviors.

(J Med Internet Res 2021;23(5):e27059) doi:10.2196/27059

KEYWORDS

Twitter; social media; human behavior; infectious disease; COVID-19; coronavirus; infodemiology; infoveillance; social distancing; shelter-in-place; mobility; COVID-19 intervention

XSL•FO
RenderX

Introduction

Health authorities can minimize the impact of an emergent infectious disease through effective and timely risk communication, vaccines and antiviral therapies, and the promotion of health behaviors, such as social distancing and personal hygiene practices [1-4]. Of these, official communication is the earliest available strategy, and its effectiveness will build trust and adherence to the remaining measures [1]. During the H1N1 influenza pandemic in 2009, most countries focused on the promotion of health behaviors [2] such as mask-wearing, avoidance of crowds, and increased disinfection after observing that such protocols contributed substantially to reduced transmission and ultimate control of disease during the SARS outbreak in 2003 [5]. Health authorities have paid less attention to the psychological factors associated with a pandemic [3,4], though such factors play a vital role in subsequent adherence to health behaviors and vaccine uptake [1]. During the emergence of the Zika virus in 2016, public health guidelines focused on preventing sexual transmission by using condoms, avoiding travel to locations with active Zika transmission, and mosquito control [6], with varying levels of compliance [7,8].

Research into the use of social media and internet data for health surveillance is a growing field. Individuals discuss a wide variety of health concerns and health behaviors online, from symptom searching [9] and personal experiences with infectious diseases [2] to dieting [10] and electronic cigarette use [11]. These data have been used to identify prominent points of discussion in relation to health topics [12-14], which can point toward more effective health policies and interventions. In addition, social media and internet data reflect temporal and spatial patterns in health behavior [9-12,15]. The association between internet data and health behavior, topics, and attitudes relevant to the public provides insight into the manner in which individuals receive health information and how that information may translate into behavioral change. Specifically for disease outbreaks, internet and social media data provide opportunities for public health officials to monitor prevalent attitudes and behaviors at a given time to target further interventions and policies.

In this work, we used social media data to better understand human behaviors relevant to COVID-19 transmission and the perceived impacts of COVID-19 on individuals. We developed a coding schema for 6 categories and 11 subcategories, which included both a wide number of behaviors as well as codes focused on the impacts of the pandemic (eg, economic and mental health impacts). We applied this schema to approximately 7200 tweets and developed supervised learning classifiers for classes with sufficient labels. We then applied these classifiers to an extensive Twitter data set and showed patterns in human behaviors temporally and spatially across the United States.

We specifically focused on the following research questions:

1. Research Question 1: What behaviors related to COVID-19 are discussed on social media websites, specifically Twitter? Using content analysis techniques similar to other social media studies (eg, Ramanadhan et al [16] and Carrotte et al [17]), we identified behaviors discussed on Twitter that could be relevant to disease transmission or the downstream impacts of COVID-19. At the outset, we were particularly interested in social distancing, hygiene, and personal protective equipment practices, but we were also interested in identifying the breadth of behaviors that might be discussed.
2. Research Question 2: How do patterns in behaviors change geospatially and temporally in the United States? Using labeled data from Research Question 1, we built classification models to identify behaviors in the larger Twitter corpus. We were interested in temporal and geospatial trends in these classified data with the goal of observing regional patterns and temporal changes that occurred in conjunction with real-world events. Prior work has used similar methods to observe patterns during Zika emergence in 2016 [15].
3. Research Question 3: How do these trends compare to other data streams, like mobility data sets? Prior work has shown that social media data are biased in multiple ways [18,19]. One way to validate our findings is to compare results using social media data to other data sources that have been useful to measure human behavior during the COVID-19 pandemic. In particular, several studies have shown that mobility data sets that rely on mobile devices (eg, smartphones) have been useful at accurately gauging reduced mobility [20,21].

Methods

Data

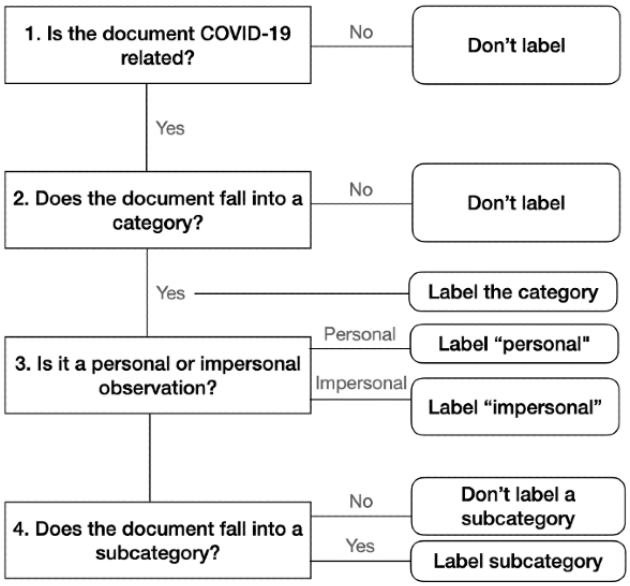
For this work, we used a data set of tweets provided by Chen et al [22]. Data collection started on January 28, 2020, and used Twitter's search application programming interface (API) to get historical tweets as early as January 21, 2020. They started with 14 keywords related to the novel coronavirus, and later expanded both keywords and individual accounts tracked over time. The data relied on Twitter's streaming API, and are thus a 1% sample of tweets that include the keywords. The original repository contained about 270 million tweets as of mid-July 2020 [22]. Of these, we were able to collect 84% (N=228,030,309).

Schema Development

The coding schema was developed by three of the authors (AD, DG, and CS) through iterative analysis of random samples of tweets from our corpus. We started initially with categories of interest (eg, social distancing and personal protective equipment) and added both categories and subcategories as they were identified in tweets, similar to prior work [16,17]. The final schema is hierarchical, where annotators can label categories and, if applicable, subcategories within the category of interest (Figure 1).

XSL•FO
RenderX

Figure 1. Decision tree for labeling.



Personal and impersonal viewpoints were labeled separately from the tweet category. Here, a *personal viewpoint* is a tweet that describes a direct observation of the behavior, meaning the individual tweeting talks about their own behavior, or a person or event that the user can directly observe. For example, the tweet "I am wearing a mask when I go out" is a personal mention of personal protective equipment, specifically mask-wearing. An *impersonal viewpoint*, in contrast, includes actions like sharing articles, retweeting, or expressing an opinion without providing evidence that the user themself engages in the behavior (eg, "Ugh, I wish more people wore masks"). This definition is the same as prior work [15]. Of note, tweets were only labeled as personal or impersonal if they were already labeled with a category. Tweets that were outside the labels of interest were not labeled for viewpoint.

Training Data

Training Annotators and Annotation

To create our training data set, 7278 tweets were selected at random from the English tweets we collected between January and May 2020, as labeling commenced in May. Using the above schema, we then trained three additional annotators. Annotators were trained using the following steps. First, a member of the team (AD) met with each individual prospective annotator and thoroughly described the schema. The prospective annotator and AD first labeled 16 example tweets together using tweets already labeled during schema development. The annotator then individually coded 160 additional tweets previously labeled by the authors. If agreement was sufficiently high (>0.6), the annotator was then given their own section of training data to code. Each tweet in our training data set was coded by two such annotators. All annotators met weekly to discuss questions about labels. All tweets with disagreements were resolved by a third annotator or via group discussion. The workflow to label tweets is given in Figure 1. Tweets can be labeled with more than one label, as applicable.

Annotator Agreement

Annotator agreement varied. Personal and impersonal labels had agreements of 0.41, 0.44, and 0.41 between the three pairs of annotators. Category-level labels had agreements of 0.77, 0.82, and 0.82, and subcategory-level labels had agreements of 0.61, 0.65, and 0.66. Distinguishing between personal and impersonal tweets was the hardest classification task because it is inherently difficult to correctly identify voice in the span of 280 characters, especially without additional context. Prior work has relied on the use of personal pronouns (eg, "I," "we," and "our") to identify personal tweets [15], but it is clear that this method has a high false negative rate because of linguistic patterns like pronoun-drop (eg, the tweet "Went to the store today and nobody was wearing masks" drops the pronoun "I" and leaves it implied) [23]. Thus, despite the difficulty in labeling these tweets, we believe it is preferable to automated methods.

Classification Algorithms

Tweet Preprocessing

Tweet URLs and usernames (@-mentions) were replaced with the tokens "URL" and "USERNAME," respectively. Consecutive characters were truncated (eg, "greaaaaaat" was truncated to "great") and punctuation was removed. Of the training data, 15% were reserved as the test set. Tweets were split using stratified sampling based on the category labels to preserve label proportions. Because of the small number of labels in several categories (Table 1), we only attempted to make classifiers for the following: personal or impersonal, social distancing (category), shelter-in-place (subcategory), monitoring (category), hygiene (category), and personal protective equipment (category).

Because personal and impersonal labels were only assigned to tweets if they fell into a category, the training data for this classifier were only those tweets with an initial label. In contrast, all other classifiers used binary classification and included all tweets that did not include the label of interest, including tweets

XSL•FO
RenderX

with no labels. As such, all classification models were built using extremely disproportionate label distributions.

Logistic Regression

Logistic regression models were implemented in Python, version 3.7.7 (Python Software Foundation), using scikit-learn [24] and the elastic net penalty. Features included all unigrams, bigrams, and trigrams of tweet text. To optimize models, grid search was used with all possible combinations of the following parameters: the elastic net penalty varied the L1 ratio from 0 (equivalent to only "L2" penalty) to 1 (equivalent to only "L1" penalty), regularization strength varied in order of magnitude from 0.001 to 1000, and chi-square feature selection was varied from 10% to 100% of the features (ie, no feature selection), in steps of 10%, to explore the impact of feature reduction on model performance.

Random Forest

Random forest models were implemented using scikit-learn's random forest classifier [24]. As in logistic regression, features included all unigrams, bigrams, and trigrams of tweet text. Again, grid search was used to optimize models. The minimum number of samples per leaf node was varied from 2 to 11 (in steps of 3), the minimum number of samples required to split an internal node ranged from 2 to 52 (in steps of 10), and the number of trees per forest was either 50 or 100. Last, we additionally varied the number of features. Because of the larger number of parameters tested here, we tried feature selections of 25%, 50%, 75%, or 100% of features (ie, no feature selection).

Classification and Bias Adjustments

Both types of models performed poorly for classifying monitoring, personal protective equipment, and hygiene. As such, we did not use these models for downstream analysis. Rather, we focus on the personal or impersonal model, the social distancing classifier, and the shelter-in-place classifier.

Though random forest models sometimes produced slightly higher F1 scores, we used the logistic regression models for overall classification and downstream analysis because of the slightly higher precision values. Said another way, in this context, we preferred fewer false positives to slightly more false negatives because we were trying to identify a particular behavior and wanted as few erroneous predictions included in the classifier as possible.

To combat the bias inherent in our classifiers, as it is clear that misclassification will occur, we used the method suggested by Daughton and Paul [25] to create confidence intervals that account for classifier error. The basic principle is to use bootstrapping to generate many samples and to subsequently weight individual classifications by the positive predictive value or negative predictive value of the classifier. The bootstrapped samples are then used to generate a 95% confidence interval around the point estimate (see Daughton and Paul [25] for full details). This method has been successfully applied in similar work focused on identifying travel change behaviors in response to Zika [15]. For this work, we used 100 bootstrapped samples to generate daily confidence intervals.

Geospatial Analysis and Comparison to Mobility and COVID-19 Data

We compared the results of our classifiers to mobility data from Descartes Labs—available at Descartes Labs [26] and described in Warren and Skillman [27]—to provide a ground truth measurement of social distancing, and to the number of confirmed COVID-19 cases in each state, as tracked by The New York Times [28]. The mobility data used geolocation services from mobile devices (eg, smartphones) to generate aggregate estimates about mobility within specific geographic areas. Descartes Labs provides data at *admin level 1* (state) mobility and *admin level 2* (county) mobility [26]. For this work, we only consider state mobility. Descartes Labs uses a normalized value of the median maximum distance traveled each day: the *m50 index*. Here, data are normalized using the median mobility per state between February 27 and March 7 (ie, a *pre–COVID-19* window). For this work, we looked at the percent change in mobility (m50 index – 100) [27], which can be interpreted as the percent change in mobility relative to the baseline period.

We used these data as a ground truth data set to validate social media tweets about social distancing and sheltering-in-place. For these comparisons, we restricted our data to those with geolocation services enabled (ie, those that used the tweet "place" to determine location), which we then aggregated by state. Here, data were aggregated to weekly data, and any weeks with fewer than 50 tweets were removed. States with fewer than 10 data points were excluded from visualization.

Results

Content Analysis and Labels

In total, 7278 tweets were read and labeled. Of these, 2202 tweets fell into the categories shown in Table 1. For each category and subcategory, the definition and an example anonymized tweet is shown. The most prevalent category by far was tweets about social distancing. Of these tweets, the vast majority were about sheltering-in-place, writ broadly, including tweets about adjusting to life at home (eg, work or school from home); tweets about entertainment, including hobbies and recipes; tweets about plans that were canceled (parties, weddings, etc); and a few tweets about a supposed "coronaboomer" phenomenon, where some suggested that the additional time spent at home would lead to an increase in babies born in 2021. In addition, we identified 53 tweets related to the mental health impacts of social distancing, including tweets about tactics to maintain positive mental health, as well as tweets describing the mental health difficulties associated with social distancing.

In other categories, we again saw a wide variety of health topics discussed. This included tweets about monitoring, of which roughly a third were about access to or experiences with COVID-19 testing; hygiene, including handwashing and cleaning protocols; and a few tweets (n=49) weighing in on COVID-19 vaccine development. Last, we also saw instances of tweets about the economic impacts of COVID-19, including on the supply chain and in terms of unemployment.

XSL•FO
RenderX

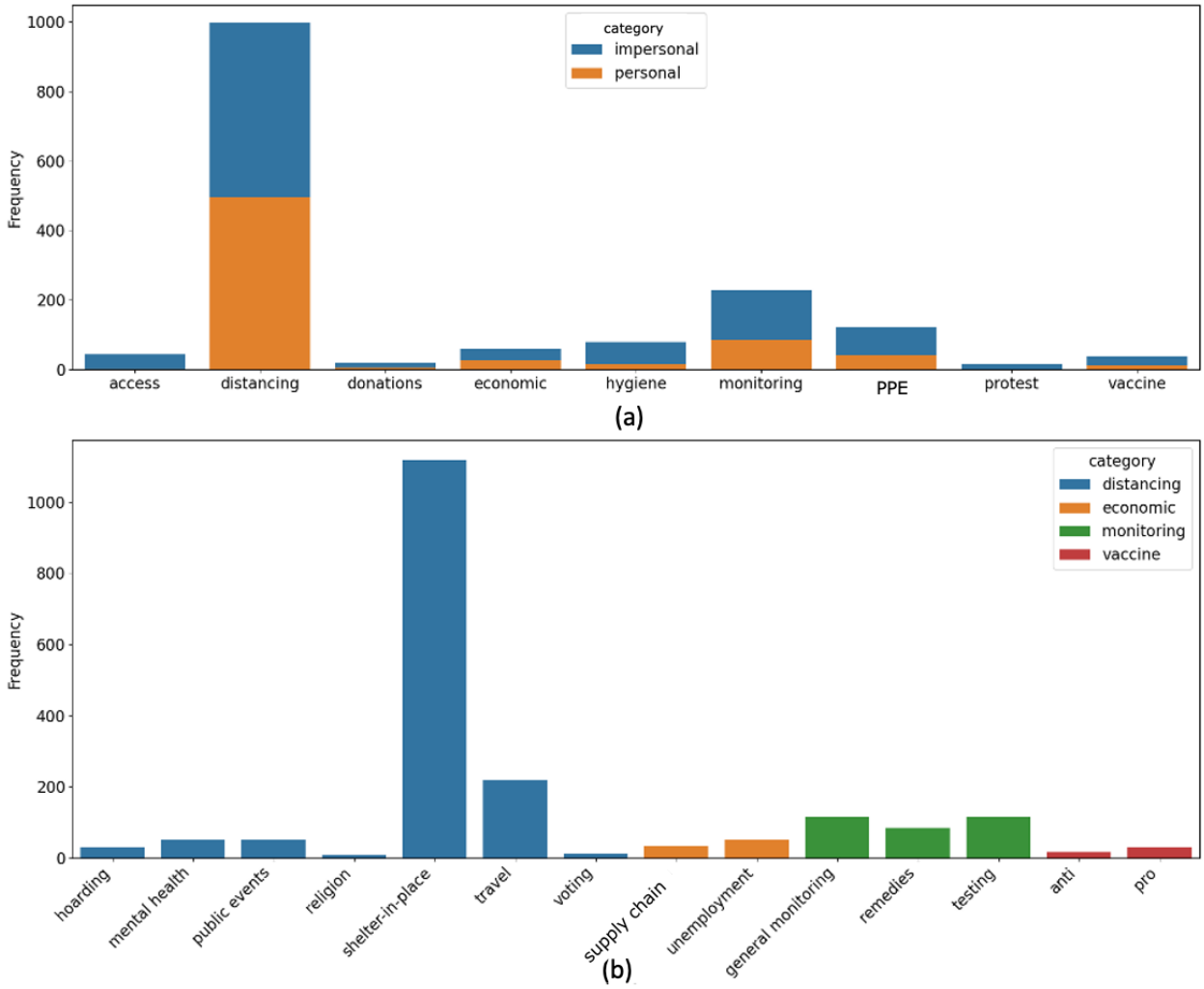
Table 1. Tweet content and relative proportion.

Category and subcategories	Definition	Example tweet (anonymized)	Tweets (N=7278), n (%)
Social distancing			
All subcategories	Discusses social distancing in either a positive or a negative way (eg, not physically seeing friends and family, not going to work, or discussing reasons why lockdowns are unnecessary)	"COVID-19 SUCKS! I can't see my family and I really miss them."	1494 (20.5)
Shelter-in-place	Discusses any aspects of shelter-in-place or stay-at-home policies; includes school or daycare (or homeschool), remote work, things to do to keep busy while staying home (eg, hobbies and recipes), canceled plans, delivery services (to avoid going out in public), and the supposed phenomenon that birth rates would increase after the pandemic ("coronaboomers")	"State going into lockdown tomorrow. I can work from home but I'm also going to catch up on my backlog Steam library!"	1117 (15.3)
Mental health	Discussions about mental health; includes suggestions of activities to maintain mental health while sheltering-in-place and documents about the mental health difficulties associated with COVID-19 and social distancing	"I'm so stressed I'm going to cry. I don't want to be where I am now, I just want to be alone for quarantine."	53 (0.7)
Voting	Decisions around voting by mail (eg, for COVID-19–related safety reasons or the opposite opinion)	"Record high cases in the past few days. It's been two weeks since the election."	12 (0.2)
Hoarding	Storing things like food, medicines, and disaster supplies	"Got a bunch of masks and gloves in case the coronavirus becomes a big deal here."	31 (0.4)
Public events	Descriptions of going to public places and choosing to not socially distance	"Airport security was super fast -- no lines at bag check."	53 (0.7)
Monitoring			
All subcategories	Behavior monitoring for illness; includes monitoring friends or family that have the disease	"I keep coming across people with sore throats and cold symptoms today. Hope it's not COVID!"	315 (4.3)
Testing	Ability or inability to get tested for COVID-19 infection; includes tweets expressing desires for improvements and increases in testing and novel testing strategies (eg, drive-through testing centers), or in combination with other tactics like contact tracing	"The complete failure in testing ramp up is really disappointing."	116 (1.6)
Remedies	Unproven treatments, advice, and/or ways to "prevent" or "cure" the disease using natural methods (eg, vitamin D)	"Anti Neo Plastons is the natural cure for Coronavirus and your body makes them naturally!"	84 (1.2)
Hygiene: all subcategories	Trying to prevent sickness by using good hygiene, including handwashing, cleaning and sanitation, and other cleanliness-related behaviors	"Just saw a kid about to use the water fountain. Their parent grabbed them and said 'NOOOOOOOO… there could be COVID!'"	94 (1.3)
Personal protective equipment: all subcategories	Using personal protective equipment to prevent illness; includes masks and gloves	"1.) Wear your mask 2.) Social distance 3.) Wash your hands! We can do this!"	164 (2.3)
Vaccine			
Provaccine	Tweets that are positive and supportive of vaccine efforts	"The work on the COVID vaccine is amazing. I can't wait to get it!"	31 (0.4)
Antivaccine	Tweets that use vaccine-averse rhetoric to describe why a vaccine will be unsafe or ill-advised	"I hope you're not in favor of the Gates vaccine. I'm not going to be tracked by a microchip!"	18 (0.2)
Economic			
Supply chain	Information or commentary about supply chain–related issues; includes information about "price gouging"	"Can we trust the food supply chain? Should we start growing our own fruits and vegetables?"	33 (0.5)
Unemployment	Includes descriptions of applying for unemployment benefits or commentary on the process; includes stimulus checks or commentary about unemployment or underemployment due to COVID-19	"I'm a driver for Uber, but I was put on medical leave after COVID-19 exposure & haven't made any money since."	53 (0.7)

XSL•FO
RenderX

A breakdown of categories by personal and impersonal labels is shown in Figure 2 (a), and subcategories are shown in Figure 2 (b). Overall, a small fraction of tweets were personal mentions; the majority of tweets were impersonal mentions related to each category (eg, mentions of articles or general opinions and suggestions that do not describe a personal behavior). This is consistent with prior work, which has found that personal mentions of health-related behavior on social media are rare [19].

Figure 2. Category distribution. Tweets are broken down by frequency of personal and impersonal labels (a) and by subcategory grouped by category (b). Categories without subcategories are not shown in (b). Only categories with at least 80 labels, and subcategories with at least 50 labels, are shown. PPE: personal protective equipment.



Because there were so few tweets in most categories, it was not feasible to build robust classifiers for most categories or subcategories. For this work, we selected for classification only the personal and impersonal classification task; the categories of social distancing, monitoring, hygiene, and personal protective equipment; and the subcategory shelter-in-place. In general, we found similar performances between random forest and logistic regression (Table 2). The exception to this trend was in the personal protective category, where the logistic regression model substantially outperformed the random forest.

For subsequent analysis, we focused on categories that achieved an F1 of at least 0.6: personal or impersonal, social distancing, and the shelter-in-place classifiers. We then applied the logistic regression models to the remaining data in our corpus of over 228 million tweets through July 2020.

XSL•FO
RenderX

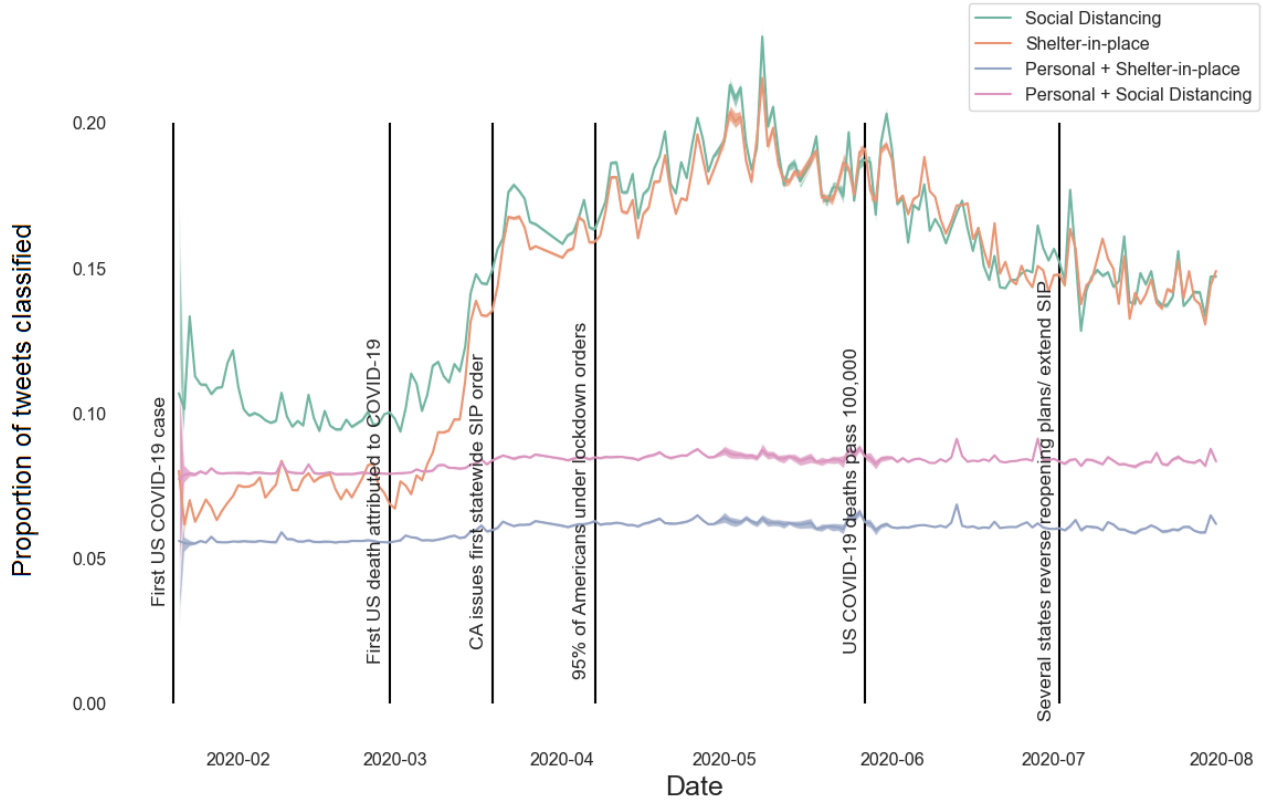
Table 2. Tweet classification results.

Classifier	Logistic regression			Random forest		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Personal or impersonal	0.76	0.50	0.60	0.72	0.57	0.64
Social distancing classifiers						
Social distancing (category)	0.73	0.59	0.66	0.71	0.61	0.66
Shelter-in-place (subcategory)	0.69	0.60	0.64	0.65	0.65	0.65
Monitoring classifiers	0.72	0.17	0.28	0.32	0.13	0.18
Hygiene classifiers	0.50	0.29	0.36	0.33	0.21	0.26
Personal protective equipment (eg, masks and gloves) classifiers	0.59	0.52	0.55	0.40	0.24	0.30

Temporal Patterns

Using the full classified corpus, we compared temporal patterns in social distancing tweets, shelter-in-place tweets, and the subsets of those groupings which were also classified as personal mentions, to important real-world events that occurred during the outbreak (Figure 3). Importantly, the proportion of tweets classified as social distancing and shelter-in-place tweets followed a predictable pattern with respect to real-world events occurring during the outbreak. Social distancing tweets occurred soon after the initial US COVID-19 case as people started to discuss initial reactions to the new disease. As states began to institute shelter-in-place orders—with California leading in late March 2020 [29]—the number of tweets about social distancing and sheltering-in-place doubled. Tweets in this category stayed high throughout the summer, as a large number of Americans were under shelter-in-place orders [29]. In early April 2020, estimates of the number of Americans told to stay at home were around 95%, despite widespread variation in how stay-at-home orders were implemented [30]. As expected, the number of personal tweets was a small fraction of the social distancing and shelter-in-place tweets more broadly. There was little variation in the temporal patterns of personal tweets; all signals came from the broader set of both personal and impersonal tweets.

Figure 3. Temporal patterns in social distancing and shelter-in-place tweets. The proportion of tweets classified as general social distancing, shelter-in-place, personal shelter-in-place, and personal social distancing are shown by date. Relevant events in the outbreak are shown as vertical lines. As states increased shelter-in-place and lockdown orders, the number of tweets about social distancing and sheltering-in-place dramatically increased. Shading shows the 95% CI calculated using classifier-adjusted bootstrapped sampling while the median is a solid line. CIs are extremely small at several time points. CA: California; SIP: shelter-in-place.



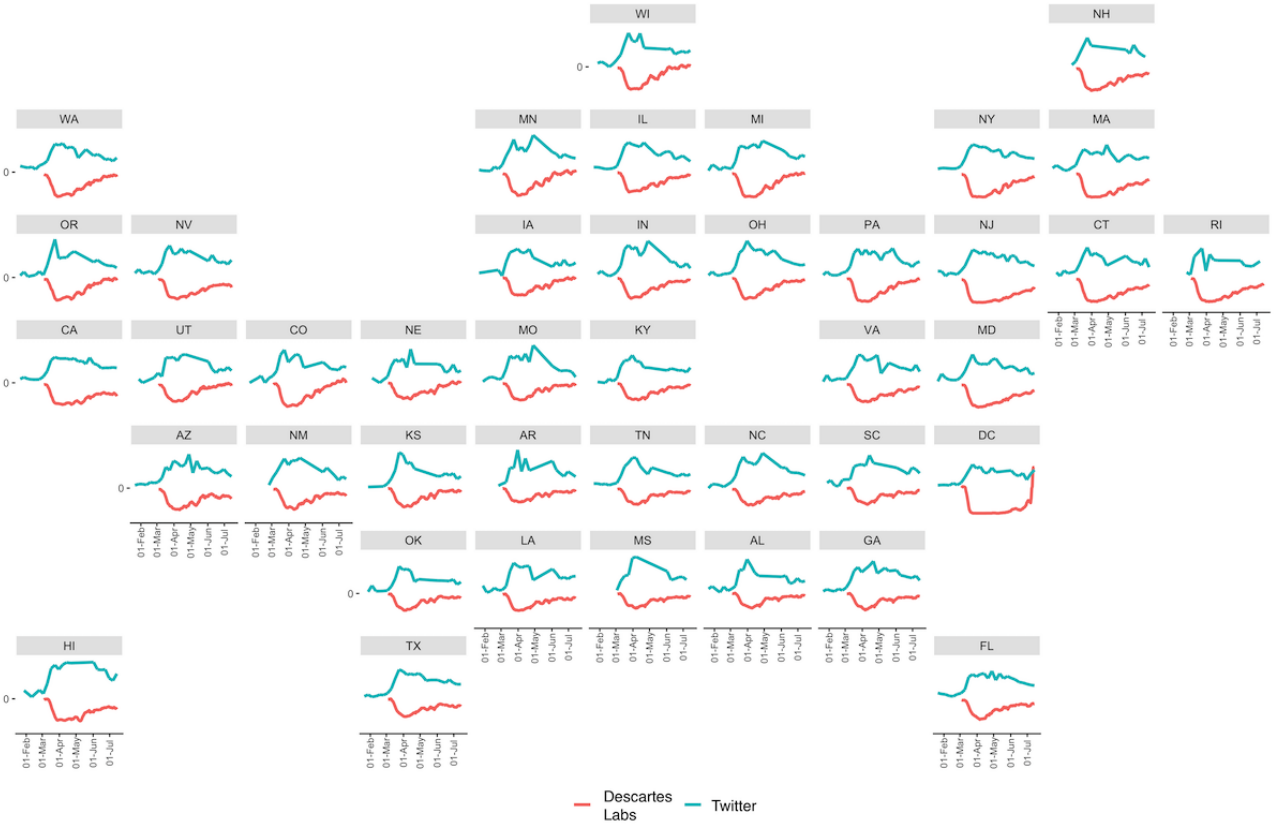
XSL•FO
RenderX

State Patterns: Comparisons to Mobility Data

To evaluate temporal patterns more closely, we considered patterns in individual states and compared them to mobility data derived from mobile phone devices (Figures 4-6) and the actual number of confirmed COVID-19 cases (Figure 6). At a high level, it is clear that there is an inverse relationship between the proportion of tweets about social distancing and the actual movement of individuals (Figure 4), indicating that social distancing conversations on Twitter may actually be reflective of real-world behavior. However, we can also see interesting regional patterns among states. For example, some of the earliest-hit states (eg, California, Washington, and New York) showed peaks in the number of tweets about social distancing in late March 2020 compared to states that saw comparatively few cases early on (eg, Florida and Georgia, which had peaks in the number of social distancing tweets in late April 2020).

Most states observed the lowest mobility in April 2020, as seen in Figure 5 (a). The day with the highest fraction of social distancing tweets was most often in March 2020, though many states observed this in April as well, as seen in Figure 5 (b). In general, most states observed these dates within ±20 days of each other, with the majority of states observing the day of minimum mobility before the day with the most tweets about social distancing, as seen in Figure 5 (c). Further, there is a strong negative correlation between the mobility data and the classified Twitter data (Figure 6). Though patterns vary by state, the average correlation is –0.42. Some states show a notably weaker signal (eg, Arkansas, New Mexico, and Rhode Island), which could be caused in part by the relative lack of data in these states. Taken together, these suggest a reasonably strong relationship between our classified Twitter data set and the ground truth mobility data. These patterns are not as clearly reflected in the relationship to confirmed COVID-19 cases. The average correlation between the proportion of tweets about social distancing and the number of confirmed COVID-19 cases is –0.08, though the strongest, which comes from Alabama, is –0.53. This suggests that, while social distancing discussions on social media are reflective of actual social distancing practices as measured by mobility data, the link to COVID-19 transmission is likely more complicated.

Figure 4. US state patterns in mobility compared to social distancing tweets from January to July 2020. Descartes Lab data showing a rolling 7-day average of percent change in mobility (divided by 5, to improve visualization) is plotted alongside the proportion of social distancing tweets per week. Both temporal and regional patterns are clear. Further, as the proportion of social distancing tweets increased, mobility measured by Descartes Labs decreased. States without sufficient Twitter data were removed from the grid. 2-letter abbreviations are used for each state.



XSL•FO
RenderX

Figure 5. Comparison of peak social distancing tweet proportions and minimum mobility. To validate our social media findings, we compared them to mobility data provided by Descartes Labs. Dates of minimum mobility are aggregated by month (a), while dates of highest proportion of tweets about social distancing, aggregated by month, are shown in (b). The difference, in days, between the date of minimum mobility and the date of highest proportion of social distancing tweets (c) show that most states observed both peaks within 20 days of one another.

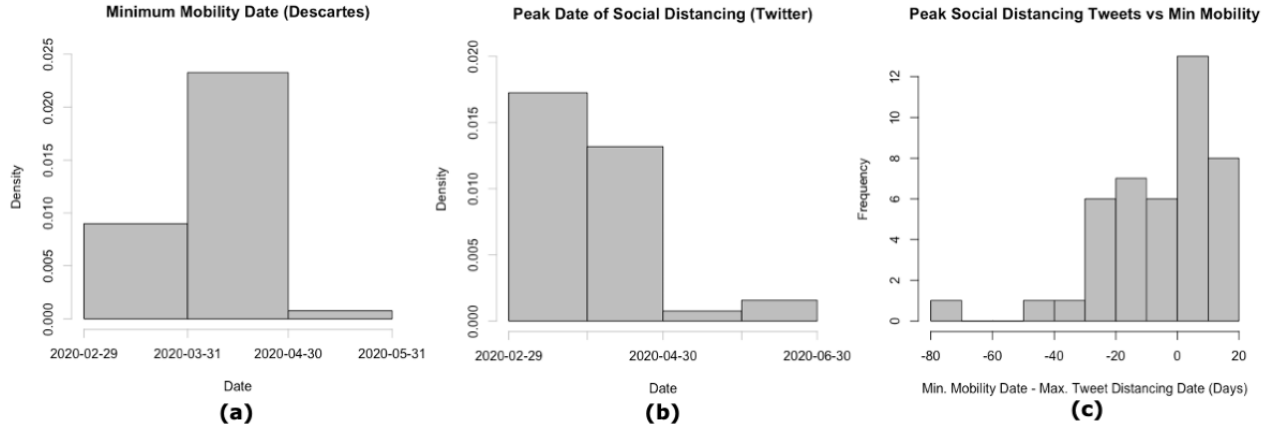
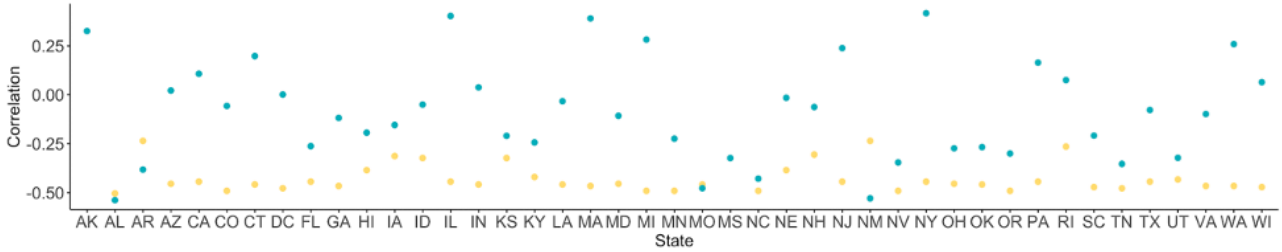


Figure 6. Correlation between confirmed COVID-19 cases or mobility and proportion of tweets about social distancing by US state. Most states have a moderate negative correlation between the proportion of tweets about social distancing and mobility data (yellow), indicating good agreement in the two signals. Some states have notably weaker negative correlations (eg, AR, NM, and RI), which could be the result of less Twitter data. Correlations between the number of confirmed COVID-19 cases and the proportion of tweets about social distancing are weak (blue), with a few notable exceptions (eg, AL). 2-letter abbreviations are used for each state.



Discussion

Principal Findings

The ongoing COVID-19 outbreak clearly illustrates the need for real-time information gathering to assess evolving beliefs and behaviors that directly impact disease spread. Historically, such information would be gathered using survey methods [5,7,31], which are time-consuming, expensive, and typically lack the ability to measure temporal and spatial variation [32]. One proposed partial solution is to use internet data (eg, search query patterns and social media data), which have been shown to correspond to disease incidence in emergent infectious disease outbreaks [23,33-35], individual risk perception [1,36,37], and risk communication [38], and have been used to identify specific health behaviors [15]. During the early stages of the current COVID-19 pandemic, social media data have been used to monitor the top concerns of individuals [39,40], characterize COVID-19 awareness [41], compare social connectedness and COVID-19 hot spots [42], monitor misinformation [40,43-45], and rapidly disseminate information [46]. Last, social media has been used as an information gathering platform during periods of uncertain information. Disease emergence is a context wherein disease risks, transmission, and treatment may be largely unclear [46]. With this context in mind, we address our findings with respect to each research question below.

What behaviors related to COVID-19 are discussed on social media websites, like Twitter? We find that there are a wide variety of behaviors discussed on social media, including mask-wearing, hygiene (eg, handwashing), testing availability and experiences, and social distancing practices. Prior work has found evidence that mask-wearing and limited mobility were behaviors adopted to reduce disease spread during SARS [5] and that handwashing would be commonly implemented by individuals during a hypothetical pandemic influenza [47]. This prior work, however, has relied on surveys to obtain data about the behaviors that individuals implement. The use of social media to complement such work would improve both the richness and the temporal and geographic scope of the data available.

Some of the identified tweets show evidence of sensitive topics. For example, we found 53 tweets related to individuals' mental health. Prior research has found that social media can be used to identify individuals with a variety of mental health concerns, including depression [48] and suicide [14]. As there is considerable work emerging about the substantial mental health impacts of COVID-19 (eg, increases in domestic violence [49] as well as depression and anxiety [50]), this could prove to be an important avenue for future work in this field.

Last, we found a small number of tweets (n=49) about vaccination related to COVID-19, of which roughly a third (n=18) showed a negative attitude. Importantly, this study was

XSL•FO
RenderX

conducted *prior* to the authorization of any vaccines in the United States. All of the tweets considered here discuss either vaccine development or a hypothetical COVID-19 vaccine. Prior research has found similarly negative tweets during the emergence of Zika [51] and the H1N1 influenza pandemic [52]. Future work analyzing these data could provide additional insight into specific reasons that populations may be hesitant to receive the COVID-19 vaccine and could inform targeted public health messaging.

How do patterns in behaviors change geospatially and temporally in the United States? As expected, the patterns in tweets classified as social distancing and shelter-in-place followed extremely similar trends. These patterns corresponded to important real-world events during the outbreak, suggesting that individuals were responding to actual events and some were describing their own personal behavior. We found, however, that tweets classified as personal mentions represented a very small subset of social distancing and shelter-in-place tweets. This is not unexpected, given that prior work has shown that personal mentions of health may be extremely uncommon [20].

How do these trends compare to other data streams, like mobility data sets, that have also shown promise in COVID-19 modeling efforts? Despite the lack of a temporal signal in tweets labeled as personal and social distancing, there was a stronger signal when comparing classified data to Descartes Labs' mobility data. We observed meaningful regional differences between states and saw that, in general, the peak number of tweets about social distancing happened within a few weeks of the actual measured minimum in mobility. This suggests that social media data may be used as a proxy for sensor data in appropriately data-rich contexts. Recent work using geotagged Twitter data to create social networks and analyze social distancing in the context of policy decisions found similar relationships and supports this finding [53].

Limitations

There are a number of limitations to consider in this work. The first is that, as mentioned above, it is known that social media data are biased in a number of ways, including demographically, and that bias differs by geographic areas [18]. Further, personal mentions of health-related information on Twitter are rare [19]. These are known limitations of using internet data and could potentially explain the variations in correlation we observed between social distancing posts and actual mobility data. Importantly, however, it is difficult to assess this without extensive prospective surveys conducted at the same time as tweet collection.

Our observed wide range in correlations between the proportion of social distancing tweets and actual COVID-19 cases in individual states is an example of the ecological fallacy. State-level COVID-19 cases represent an aggregate measure of a state's behavior, while tweets represent individual actions and observations. The available data do not allow us to probe the reasons for the variation, but a number of possible factors could be at play. Individuals' social distancing thoughts at a specific moment in time will be influenced by contextual information about other aspects of their lives. For example, people that tweet in support of social distancing may have in-person jobs or be in high-risk groups, which could motivate them to use social media platforms to voice support for public health measures. The stronger correlation with mobility outcomes is expected by this same argument because mobility is more directly representative of individual actions.

Additionally, tweeting norms could be systematically different across the country (eg, people in different states might be more or less likely to talk about social distancing based on the policies in place and the perceived threat of COVID-19). It is also possible that there are differences in which individuals use Twitter and have geolocation services enabled in different states. In an operational context, it is hugely important to combine internet data with traditional data streams in order to provide a more complete picture of an evolving scenario. Future work should focus on targeted studies to better understand potential bias.

An additional known source of bias comes from imperfect classification. Our classifiers performed similarly to other classifiers used to identify health behaviors [15], but were clearly not perfect. To account for known classifier bias, we used an adjusted bootstrapping method from Daughton and Paul [25], which generates accurate confidence intervals despite classifier error.

We validated our work using mobility data from Descartes Labs. However, there are a number of mobility data sources available [54]. Prior work indicates that these data have similar patterns [54], but it is possible that using a different source would produce slightly different validation results.

Conclusions

Behavior changes and policy decisions that occur early within an outbreak have the largest effects on disease dynamics [55,56]. Real-time conversations about health behaviors, in addition to other behavioral data sources such as mobility metrics or media consumption (eg, home television viewing [55]), could help improve overall knowledge and policy decisions in the early stages of an epidemic and could better capture dynamic changes caused by uncoordinated behavioral change. Using such data has the unique capability to inform public health decisions as an outbreak emerges, especially with respect to public health communication. The World Health Organization suggests a communication checklist to prepare for and minimize morbidity and mortality in the event of a pandemic [57,58]. The checklist emphasizes building public trust through early communication, even with incomplete information, and evaluating the impact of communication programs to assess whether recommendations are being followed. The use of social media streams as a simultaneous real-time measure of public sentiment toward messaging and a dynamic evaluation tool of communication effectiveness could be invaluable in minimizing effects from a future disease outbreak.

XSL•FO
RenderX

Acknowledgments

ARD, CDS, and DG created the labeling schema. DG, NP, TP, ARD, CWR, GF, and NYVC collected and analyzed the Twitter data. ARD, CDS, DG, IC, GN, and NM labeled the tweets. ARD and CWR built the supervised learning models, and ARD implemented the classifier-adjusted bootstrapped sampling. MB collected the mobility data and created several figures. ARD, CDS, and MB wrote the initial paper. All authors provided critical revisions to the paper. ARD led the project. Research support was provided by the Laboratory Directed Research and Development program of Los Alamos National Laboratory (project No. 20200721ER) and the US Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy (Contract No. 89233218CNA000001). The Los Alamos National Laboratory Review & Approval System reporting number is LA-UR-21-20074.

Conflicts of Interest

None declared.

References

1. Taylor S. The Psychology of Pandemics: Preparing for the Next Global Outbreak of Infectious Disease. Newcastle upon Tyne, UK: Cambridge Scholars Publishing; 2019.
2. Bults M, Beaujean DJ, Richardus JH, Voeten HA. Perceptions and behavioral responses of the general public during the 2009 influenza A (H1N1) pandemic: A systematic review. Disaster Med Public Health Prep 2015 Apr;9(2):207-219. [doi: 10.1017/dmp.2014.160] [Medline: 25882127]
3. Douglas PK, Douglas DB, Harrigan DC, Douglas KM. Preparing for pandemic influenza and its aftermath: Mental health issues considered. Int J Emerg Ment Health 2009;11(3):137-144. [Medline: 20437844]
4. Shultz JM, Espinel Z, Flynn BW, Hoffman Y, Cohen RE. Deep Prep: All-Hazards Disaster Behavioral Health Training. Miami, FL: Miller School of Medicine, University of Miami; 2008.
5. Lau JTF, Yang X, Pang E, Tsui HY, Wong E, Wing YK. SARS-related perceptions in Hong Kong. Emerg Infect Dis 2005 Mar;11(3):417-424 [FREE Full text] [doi: 10.3201/eid1103.040675] [Medline: 15757557]
6. MacDonald PDM, Holden EW. Zika and public health: Understanding the epidemiology and information environment. Pediatrics 2018 Feb 01;141(Supplement 2):S137-S145. [doi: 10.1542/peds.2017-2038b]
7. Darrow W, Bhatt C, Rene C, Thomas L. Zika virus awareness and prevention practices among university students in Miami: Fall 2016. Health Educ Behav 2018 Dec;45(6):967-976. [doi: 10.1177/1090198118760687] [Medline: 29534620]
8. Mendoza C, Jaramillo G, Ant TH, Power GM, Jones RT, Quintero J, et al. An investigation into the knowledge, perceptions and role of personal protective technologies in Zika prevention in Colombia. PLoS Negl Trop Dis 2020 Jan;14(1):e0007970 [FREE Full text] [doi: 10.1371/journal.pntd.0007970] [Medline: 31961867]
9. White RW, Horvitz E. From health search to healthcare: Explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc 2014;21(1):49-55 [FREE Full text] [doi: 10.1136/amiajnl-2012-001473] [Medline: 23666794]
10. Coogan S, Sui Z, Raubenheimer D. Gluttony and guilt: Monthly trends in internet search query data are comparable with national-level energy intake and dieting behavior. Palgrave Commun 2018 Jan 9;4(1):1-9. [doi: 10.1057/s41599-017-0055-7]
11. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011 Apr;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007] [Medline: 21406279]
12. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
13. Paul MJ, Dredze M. You are what you tweet: Analyzing Twitter for public health. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14137/13986
14. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions - Evidence from Twitter. J Am Med Inform Assoc 2017 May 01;24(3):496-502 [FREE Full text] [doi: 10.1093/jamia/ocw133] [Medline: 27707822]
15. Daughton AR, Paul MJ. Identifying protective health behaviors on Twitter: Observational study of travel advisories and Zika virus. J Med Internet Res 2019 May 13;21(5):e13090 [FREE Full text] [doi: 10.2196/13090] [Medline: 31094347]
16. Ramanadhan S, Mendez SR, Rao M, Viswanath K. Social media use by community-based organizations conducting health promotion: A content analysis. BMC Public Health 2013 Dec 05;13:1129 [FREE Full text] [doi: 10.1186/1471-2458-13-1129] [Medline: 24313999]
17. Carrotte ER, Prichard I, Lim MSC. "Fitspiration" on social media: A content analysis of gendered images. J Med Internet Res 2017 Mar 29;19(3):e95 [FREE Full text] [doi: 10.2196/jmir.6368] [Medline: 28356239]

XSL•FO
RenderX

18. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the demographics of Twitter users. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234
19. Daughton AR, Chunara R, Paul MJ. Comparison of social media, syndromic surveillance, and microbiologic acute respiratory infection data: Observational study. JMIR Public Health Surveill 2020 Apr 24;6(2):e14986 [FREE Full text] [doi: 10.2196/14986] [Medline: 32329741]
20. Engle S, Stromme J, Zhou A. Staying at home: Mobility effects of COVID-19. SSRN J 2020:1-16 (forthcoming). [doi: 10.2139/ssrn.3565703]
21. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. Science 2020 Apr 10;368(6487):145-146. [doi: 10.1126/science.abb8021] [Medline: 32205458]
22. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
23. Lamb A, Paul MJ, Dredze M. Separating fact from fear: Tracking flu infections on Twitter. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013 Presented at: 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 9-14, 2013; Atlanta, GA p. 789-795.
24. 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 Nov;12:2825-2830 [FREE Full text] [doi: 10.5555/1953048.2078195]
25. Daughton AR, Paul MJ. Constructing accurate confidence intervals when aggregating social media data for public health monitoring. In: Proceedings of the International Workshop on Health Intelligence (W3PHAI 2019). 2019 Presented at: International Workshop on Health Intelligence (W3PHAI 2019); January 27-February 1, 2019; Honolulu, HI.
26. Data for mobility changes in response to COVID-19. GitHub. Santa Fe, NM: Descartes Labs URL: https://github.com/descarteslabs/DL-COVID-19 [accessed 2021-05-06]
27. Warren MS, Skillman SW. Mobility changes in response to COVID-19. ArXiv. Preprint posted online on March 31, 2020. [FREE Full text]
28. Coronavirus (Covid-19) data in the United States. GitHub. New York, NY: The New York Times URL: https://github.com/nytimes/covid-19-data [accessed 2021-05-06]
29. State “shelter-in-place” and “stay-at-home” orders. FINRA. URL: https://www.finra.org/rules-guidance/key-topics/covid-19/shelter-in-place [accessed 2020-12-23]
30. Mervosh S, Lu D, Swales V. See which states and cities have told residents to stay at home. The New York Times. 2020 Apr 20. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html [accessed 2021-01-05]
31. Chandrasekaran N, Marotta M, Taldone S, Curry C. Perceptions of community risk and travel during pregnancy in an area of Zika transmission. Cureus 2017 Jul 26;9(7):e1516 [FREE Full text] [doi: 10.7759/cureus.1516] [Medline: 28959511]
32. Blaikie N. Designing Social Research: The Logic of Anticipation. 2nd edition. Cambridge, UK: Polity Press; 2009.
33. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011 May;5(5):e1206. [doi: 10.1371/journal.pntd.0001206] [Medline: 21647308]
34. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the First Workshop on Social Media Analytics (SOMA '10). 2010 Presented at: First Workshop on Social Media Analytics (SOMA '10); July 25, 2010; Washington, DC p. 155-122. [doi: 10.1145/1964858.1964874]
35. Watad A, Watad S, Mahroum N, Sharif K, Amital H, Bragazzi NL, et al. Forecasting the West Nile virus in the United States: An extensive novel data streams-based time series analysis and structural equation modeling of related digital searching behavior. JMIR Public Health Surveill 2019 Feb 28;5(1):e9176 [FREE Full text] [doi: 10.2196/publichealth.9176] [Medline: 30601755]
36. Hassan MS, Halbusi HA, Najem A, Razali A, Williams KA, Mustamil NM. Impact of risk perception on trust in government and self-efficiency during COVID-19 pandemic: Does social media content help users adopt preventative measures? Research Square. Preprint posted online on July 16, 2020. [doi: 10.21203/rs.3.rs-43836/v1]
37. Oh SH, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: The mediating role of self-relevant emotions and public risk perception. Health Commun 2020 Feb 16:1-10. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]
38. Ding H, Zhang J. Social media and participatory risk communication during the H1N1 flu epidemic: A comparative study of the United States and China. China Media Res 2010;6(4):80-91 [FREE Full text]
39. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
40. Singh L, Bansal S, Bode L, Budakb C, Chic G, Kawintiranona K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. ArXiv. Preprint posted online on March 31, 2020. [FREE Full text]

XSL•FO
RenderX

41. Saad M, Hassan M, Zaffar F. Towards characterizing COVID-19 awareness on Twitter. ArXiv. Preprint posted online on May 21, 2020. [FREE Full text]
42. Bailey M, Cao R, Kuchler T, Stroebel J, Wong A. Social connectedness: Measurement, determinants, and effects. J Econ Perspect 2018 Aug 01;32(3):259-280. [doi: 10.1257/jep.32.3.259]
43. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
44. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: An analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
45. Gerts D, Shelley CD, Parikh N, Pitts T, Watson Ross C, Fairchild G, et al. "Thought I'd share first" and other conspiracy theory tweets from the COVID-19 infodemic: Exploratory study. JMIR Public Health Surveill 2021 Apr 14;7(4):e26527 [FREE Full text] [doi: 10.2196/26527] [Medline: 33764882]
46. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: Early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
47. Sadique MZ, Edmunds WJ, Smith RD, Meerding WJ, de Zwart O, Brug J, et al. Precautionary behavior in response to perceived threat of pandemic influenza. Emerg Infect Dis 2007 Sep;13(9):1307-1313 [FREE Full text] [doi: 10.3201/eid1309.070372] [Medline: 18252100]
48. De Choudhury CM, Gamon M, Counts S, Horvitz S. Predicting depression via social media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Boston, MA URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351
49. Kofman YB, Garfin DR. Home is not always a haven: The domestic violence crisis amid the COVID-19 pandemic. Psychol Trauma 2020 Aug;12(S1):S199-S201 [FREE Full text] [doi: 10.1037/tra0000866] [Medline: 32478558]
50. 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]
51. Ghenai A, Mejova Y. Catching Zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on Twitter. In: Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: 2017 IEEE International Conference on Healthcare Informatics (ICHI); August 23-26, 2017; Park City, UT p. 518. [doi: 10.1109/ichi.2017.58]
52. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
53. Porcher S, Renault T. Social distancing beliefs and human mobility: Evidence from Twitter. ArXiv. Preprint posted online on August 10, 2020. [FREE Full text] [doi: 10.1371/journal.pone.0246949]
54. Huang X, Li Z, Jiang Y, Ye X, Deng C, Zhang J, et al. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. Int J Digit Earth 2021 Feb 17;14(4):424-442. [doi: 10.1080/17538947.2021.1886358]
55. Schwarzinger M, Flicoteaux R, Cortarenoda S, Obadia Y, Moatti J. Low acceptability of A/H1N1 pandemic vaccination in French adult population: Did public health policy fuel public dissonance? PLoS One 2010 Apr 16;5(4):e10199 [FREE Full text] [doi: 10.1371/journal.pone.0010199] [Medline: 20421908]
56. Springborn M, Chowell G, MacLachlan M, Fenichel EP. Accounting for behavioral responses during a flu epidemic using home television viewing. BMC Infect Dis 2015 Jan 23;15:21 [FREE Full text] [doi: 10.1186/s12879-014-0691-0] [Medline: 25616673]
57. World Health Organization, Department of Communicable Disease Surveillance and Response. WHO Guidelines for Epidemic Preparedness and Response to Measles Outbreaks. Geneva, Switzerland: World Health Organization; 1999 May. URL: http://www.who.int/csr/resources/publications/measles/whocdscsrisr991.pdf [accessed 2016-07-27]
58. World Health Organization, Department of Communicable Disease Surveillance and Response, Global Influenza Programme. WHO Checklist for Influenza Pandemic Preparedness Planning. Geneva, Switzerland: World Health Organization; 2005. URL: https://www.who.int/influenza/resources/documents/FluCheck6web.pdf?ua=1 [accessed 2020-05-06]

Abbreviations

API: application programming interface

XSL•FO
RenderX

Edited by C Basch; submitted 11.01.21; peer-reviewed by X Zhou, L Guo, Z Jin; comments to author 01.03.21; revised version received 08.03.21; accepted 17.04.21; published 25.05.21.

Please cite as:
Daughton AR, Shelley CD, Barnard M, Gerts D, Watson Ross C, Crooker I, Nadiga G, Mukundan N, Vaquera Chavez NY, Parikh N, Pitts T, Fairchild G
Mining and Validating Social Media Data for COVID-19–Related Human Behaviors Between January and July 2020: Infodemiology Study
J Med Internet Res 2021;23(5):e27059
URL: https://www.jmir.org/2021/5/e27059
doi:10.2196/27059
PMID:33882015

©Ashlynn R Daughton, Courtney D Shelley, Martha Barnard, Dax Gerts, Chrysm Watson Ross, Isabel Crooker, Gopal Nadiga, Nilesh Mukundan, Nidia Yadira Vaquera Chavez, Nidhi Parikh, Travis Pitts, Geoffrey Fairchild. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Detecting Impending Stroke From Cognitive Traits Evident in Internet Searches: Analysis of Archival Data

Sigal Shaklai[1,2,3], MD, PhD; Ran Gilad-Bachrach[4,5], PhD; Elad Yom-Tov[6,7], PhD; Naftali Stern[1,2,3], MD

[1]Institute of Endocrinology, Metabolism and Hypertension, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

[2]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[3]Sagol Center for Epigenetics of Aging and Metabolism, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

[4]Faculty of Bio-Medical Engineering, Tel Aviv University, Tel Aviv, Israel

[5]Edmond J Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel

[6]Microsoft Research, Herzeliya, Israel

[7]Faculty of Industrial Engineering and Management, Technion, Haifa, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
3 Alan Turing
Herzeliya, 4672415
Israel
Phone: 972 779391359
Email: eladyt@yahoo.com

Abstract

Background: Cerebrovascular disease is a leading cause of mortality and disability. Common risk assessment tools for stroke are based on the Framingham equation, which relies on traditional cardiovascular risk factors to predict an acute event in the near decade. However, no tools are currently available to predict a near/impending stroke, which might alert patients at risk to seek immediate preventive action (eg, anticoagulants for atrial fibrillation, control of hypertension).

Objective: Here, we propose that an algorithm based on internet search queries can identify people at increased risk for a near stroke event.

Methods: We analyzed queries submitted to the Bing search engine by 285 people who self-identified as having undergone a stroke event and 1195 controls with regard to attributes previously shown to reflect cognitive function. Controls included random people 60 years and above, or those of similar age who queried for one of nine control conditions.

Results: The model performed well against all comparator groups with an area under the receiver operating characteristic curve of 0.985 or higher and a true positive rate (at a 1% false-positive rate) above 80% for separating patients from each of the controls. The predictive power rose as the stroke date approached and if data were acquired beginning 120 days prior to the event. Good prediction accuracy was obtained for a prospective cohort of users collected 1 year later. The most predictive attributes of the model were associated with cognitive function, including the use of common queries, repetition of queries, appearance of spelling mistakes, and number of queries per session.

Conclusions: The proposed algorithm offers a screening test for a near stroke event. After clinical validation, this algorithm may enable the administration of rapid preventive intervention. Moreover, it could be applied inexpensively, continuously, and on a large scale with the aim of reducing stroke events.

(J Med Internet Res 2021;23(5):e27084) doi:10.2196/27084

KEYWORDS

search engines; diagnosis; screening; stroke; risk; internet; trend; infodemiology; archive; prospective; algorithm

Introduction

Cerebrovascular disease is a leading cause of mortality and disability. Further, with the worldwide aging of the population, the rise in stroke mortality is only second to that attributed to heart disease [1]. The overall care of stroke survivors comprises an immense social and economic burden due to the high incidence (~30%) of residual cognitive and physical disabilities

XSL•FO
RenderX

[2]. Hypertension, atrial fibrillation, hyperlipidemia, smoking, obesity, and diabetes comprise the major treatable risk factors that underly the occurrence of most stroke events [3]. With the exception of obesity and smoking, these conditions are often asymptomatic and therefore can be missed with treatment deferred for years, unless incidentally detected, clinically suspected, or realized during an acute vascular event. Even when subjects or physicians are aware of these conditions, treatment is not consistently initiated and is often insufficient or discontinued [4,5] due to patient or health care system neglect. Effective means now exist that can substantially lower the rate of stroke shortly after their implementation, as best exemplified by the early benefits of immediate treatment with anticoagulants for atrial fibrillation and blood pressure–lowering drugs for hypertension [6,7]. Therefore, the timely detection of an impending stroke might open a time window of opportunity for immediate medical intervention that can save lives and prevent poststroke disability. To our knowledge, the only risk calculator for impending stroke has been constructed in subjects who have already experienced a previous single stroke, aiming to predict the 90-day risk of recurrent stroke [4], but not for first-ever stroke events.

Computer-based analysis of electronic health records has the potential to provide major novel insights of benefit both to specific individuals in the context of personalized medicine as well as on the level of population-wide health care and policy [8]. However, such analysis mostly relies on traditional risk factors that have already been at the core of daily practice for several decades. The most widely accepted models for stroke prediction, such as those of the American College of Cardiology/American Heart Association cardiovascular disease risk algorithm and the Framingham risk score actually offer a more general predictive target, usually referred to as an "event," which includes an overall risk of occurrence of stroke and other vascular events such as myocardial infarction, death from coronary heart disease, congestive heart failure, incident angina, or intermittent claudication [9]. Further, although these risk profiling models provide clinically critical information in that they calculate the likelihood of an event in the near decade, they have not been constructed to evaluate the risk of an impending short-term cardiovascular threat.

An alternative approach for predicting stroke relies on identification of covert cerebrovascular disease, which often precedes stroke. Covert cerebrovascular disease is associated with subtle cognitive and motor deficits, and increased risk for stroke and further cognitive decline [10,11]. Currently, this can be diagnosed by either neuroimaging or in-person assessment using conventional cognitive screening tests; however, both tools are costly, not effective on the population as a whole, and not continuous [12].

The way that an individual interacts with their computer has been suggested to represent an as-yet unexploited means of assessing everyday cognition [13]. Computer interaction requires motor, language, and executive functions; ability to operate the keyboard and mouse; comprehend and create text; plan; and focus attention. Routine use such as querying of search engines has been shown to correlate with parts of standard cognitive tests, and may thus be used as a continuous, unobtrusive, inexpensive monitoring application [13]. More broadly, internet data have been shown to be a useful source for studying health and improving public health. This has become an active field of research, known as infodemiology, a term coined by Eysenbach [14,15]. Due to their ubiquitous nature and anonymity [16], search engine queries have been shown to be useful for screening for a variety of diseases, including Alzheimer disease [17], Parkinson disease [18], several types of cancer [19-21], diabetes [22], and eating disorders [23]. Similarly, social media postings have been shown to change prior to emergency department visits [24].

In this study, we evaluated the possibility that subjects disclosing the fact that they have experienced a stroke can be singled out from subjects without such a self-declared history via their internet-based communications prior to the occurrence of the stroke. If prestroke patients show unique alterations in communication features, this could potentially be utilized as a "last minute" alarm to seek prompt medical help, with the hope that an impending stroke can be prevented.

Methods

Data

Dataset 1 comprised all queries submitted to the Bing search engine by people in the United States during a 2-year period beginning July 2017. The data comprised the query text, the time and date, an anonymous user identifier, and data on the interaction between the user and the search results page.

To identify a "patient" cohort, we found users who queried the search engine for phrases that indicated that the user underwent a stroke, such as "I had a stroke" or "I was diagnosed with a stroke" (referred to as the "reference query"). Specifically, we defined the reference queries to be queries containing one of the phrases "I had," "I experienced," or "I suffered" *and* one of the phrases "stroke," "cerebrovascular accident," "CVA," "cerebrovascular event," "transient ischemic attack," "brain infarction," "brain ischemia," and "cerebral ischemia."

Additionally, we required that users were active for at least 10 days before and after this query, made at least 50 queries, and had no more than 3 periods of inactivity of 24 hours or more in the 60 days prior to the reference query. We also required that in the 60 days prior to the reference query, the user did not issue queries for between 1 and 10 days, which, we hypothesized, represented the time that the user was hospitalized.

We chose several populations as controls, including a random sample of people aged 60-64, 65-74, or 75 years or older, and random samples of people aged 60 years or older who queried one of the following conditions (separately): depression, migraine, B12 deficiency, heart attack, hypothyroidism, surgery, atrial fibrillation, hypertension, and Alzheimer disease. Table 1 lists the terms used to identify these conditions. Ages were determined according to information provided by users at the time of registration to Bing.

The "reference query" for the control populations was chosen as a random query from their queries (for people selected by their age) or the first query mentioning the condition of interest

XSL•FO
RenderX

(for those who queried having experienced one of the control conditions above). All control populations were filtered for the same activity levels (as described above) as the patient cohort. Dataset 2 was similar to Dataset 1, but was taken from July 2019 to June 2020 and was limited to users who were aged 60 years or more.

Table 1. Number of users in the control cohorts according to various conditions and the search terms used to identify the group.

Cohort	Terms or criteria used for identification	Users, N
60-64 years old	According to information provided at registration to Bing	108
65-74 years old	According to information provided at registration to Bing	82
>75 years old	According to information provided at registration to Bing	38
Depression	depression, depressed, ssri, citalopram, celexa, escitalopram, lexapro, fluoxetine, prozac, fluvoxamine, luvox, paroxetine, paxil, sertraline, zoloft, atomoxetine, Strattera, pristiq, desvenlafaxine, cymbalta, duloxetine, levomilnacipran, fetzima, milnacipran, ixel, savella, dalcipran, toledomin, sibutramine, meridia, tramadol, ultram, venlafaxine, effexor	85
Migraine	migraine, metoprolol, Lopressor, valproate, Depakote, epilim, topiramate, Topamax	66
B12 deficiency	B12	48
Heart attack	heart attack, acute myocardial infarction	101
Hypothyroidism	hypothyroidism, levothyroxine, levoxyl, synthroid, tirosint, unithroid	59
Surgery	Prostatectomy, hip replacement, surgery	369
Atrial fibrillation	atrial fibrillation, anticoagulants, eliquis, apixaban, xarelto, rivaroxaban, pradaxa, dabigatran, lixiana, edoxaban, coumadin, warfarin, antiarrythmic, amiodarone, procor, cordarone, sotalol, flecainide, tambocor, propafenone, rythmol, dronedarone, multaq, dofetilide, tikosyn	86
Hypertension	hypertension, angiotensin receptor blocker, irbesartan, telmisartan, candesartan, atacand, valsartan, diovan, losartan, cozaar, losardex, olmesartan, benicar, ace inhibitor, benazepril, lotensin, captopril, cilazapril, enalapril, enaladex, fosinopril, lisinopril, ramipril, moexipril, cardiotensin, perindopril, quinapril, accupril, trandolapril, mavik, mineralocorticoid receptor antagonists, spironolactone, aldactone, carospir, eplerenone, inspra, diuretics, chlorthalidone, calcium channel blocker, amlodipine, norvasc, katerzia, diltiazem, cardizem, matzim, taztia, tiazac, verapamil, verap, alpha blocker, doxazosin, cardura, cadex	78
Alzheimer disease	((I or my) AND (forgetting, misplacing, can't find, forget name)) OR Alzheimer OR dementia	75

Data Analysis

Each user was represented through the average values of several attributes and the SDs thereof, computed for all queries within a 2-day moving-window period. These attributes were chosen to represent cognitive ability [13,25,26], activity routine, and indications related to risk factors for stroke. The list of attributes is as follows: (1) number of words per query; (2) time of day; (3) the likelihood of the query string in the entire population; (3) number of new words in the query, compared to all previous queries by the user; (4) time since previous session; (5) number of queries per day and per hour; (6) time from the display of the results to the first click on a result; (7) farthest link clicked by the user; (8) use of automatic spelling correction; (9) whether the query was submitted by this user in the past; and (10) specific keywords in the query, including personal references, mention of relevant symptoms, or mention of certain medicinal drugs.

A session was defined as a continuous period of querying, followed by a break of 30 minutes or longer. Additionally, we quantified the ratio or each attribute to its values for the same user 10 days before. As noted above, we assumed that the gap in user queries just before the reference query was likely due to hospitalization following a stroke. Therefore, we defined the time of the last query prior to the gap as time zero for each user and calculated the time of each user's query relative to time zero. We refer to this as the "relative time" of a query. Patterns were augmented with the average value of each attribute within the relative time of –150 to –120 days.

Unless otherwise stated, we used all data from 30 days prior to a gap in the queries, which was followed by the reference query and up to that gap, to build a random forest model with 1000 trees to distinguish the patient cohort from each of the control cohorts. Ten-fold cross-validation at the user level was applied to reduce the likelihood of overfitting. All processes were performed using Matlab 2019. This study was approved by the Institutional Review Board of Technion, Israel Institute of Technology.

Results

Separation of the Stroke Cohort From Control Cohorts

We identified 285 users in the patient cohort; Table 1 shows the number of users in each of the control cohorts. Figure 1 shows the receiver operating characteristic (ROC) curve [27] for distinguishing the patient cohort and the control by age cohorts. These data comprise only the queries prior to the gap preceding the reference query. The ROC curve showed that excellent differentiation between the patient and control cohorts is possible.

XSL•FO
RenderX

Figure 2 shows the area under the ROC curve (AUC) [27] and the true positive rate at a 1% false-positive rate for separating the patient cohort from each of the control cohorts. Very good performance was reached for all conditions. In addition, some cardiovascular diseases (heart attack, hypertension, and migraine) and certain mental diseases (depression) appeared to be harder to separate from stroke.

Figure 1. Receiver operating characteristic curve for distinguishing users who underwent stroke from controls for people of different ages.

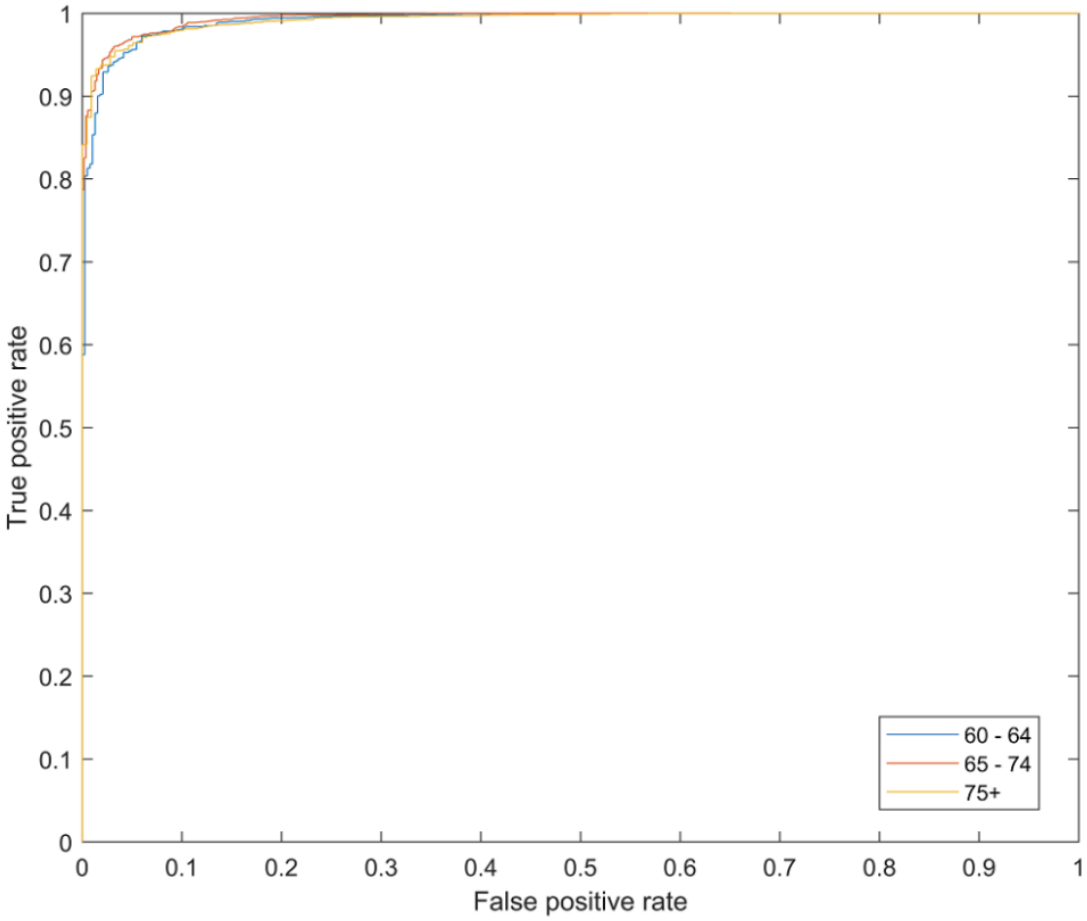
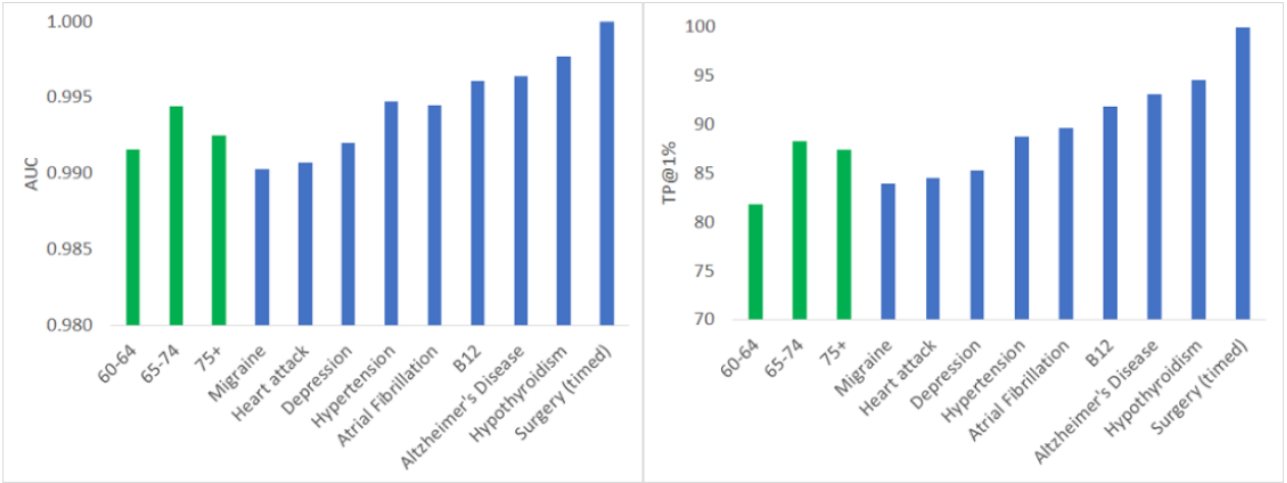


Figure 2. Area under the receiver operating characteristic curve (AUC, left) and the true positive rate at a 1% false-positive rate for separating the patient cohort from each of the control cohorts.



The best attributes for separating the stroke cohort from the other cohorts were estimated by finding the increase in prediction error if the values of that variable were permuted for the test observations, using the appropriate Matlab function (OOBPermutedPredictorDeltaError). The five most important attributes for separating the stroke cohort from each of the control conditions were (in descending order of the number of conditions in which they appeared in the top 5): average query likelihood, SD of the query likelihood, average number of queries per session, average number of spelling mistakes, and average number of repeated queries.

XSL•FO
RenderX

Although the model does not use specific vocabulary in identifying people at risk, it is nevertheless interesting to study the differences in the vocabulary used by the patient cohort and the control groups, and how the vocabulary changes after the stroke event.

Risk factors for stroke include high blood pressure, hypertension, diabetes, insulin, heart disease, cardiomyopathy, heart failure, atrial fibrillation, blood clot, smoking, cocaine, amphetamines, sickle cell disease, vasculitis, bleeding, overweight, stress, and cholesterol. Among people in the patient cohort, 32.2% queried one or more of these risk factors prior to day zero and 56.4% queried for these terms after day zero. Among people 60 years or older in the control cohort, the corresponding percentages were 4.3% and 4.6%.

Medical drugs used to treat survivors of stroke include plavix, clopidogrel, aspirin, dipyridamole, pradaxa, dabigatran, eliquis, apixaban, xarelto, rivaroxaban, atorvastatin, lipitor, simvastatin, simovil, rosuvastatin, crestor, pravastatin, and pravachol. Among people in the patient cohort, 12.5% queried one or more of these drugs prior to day zero and 28.2% queried for these terms after day zero. Among people 60 years or older in the control group, the corresponding percentages were 0.7% and 0.8%. We assume that some of the people mentioning the drugs in the control group and prior to the day of stroke in the treatment group did so because they received these drugs due to their being in high-risk groups for the disease.

These two results suggest that users in the patient cohort did indeed undergo a stroke at the estimated date.

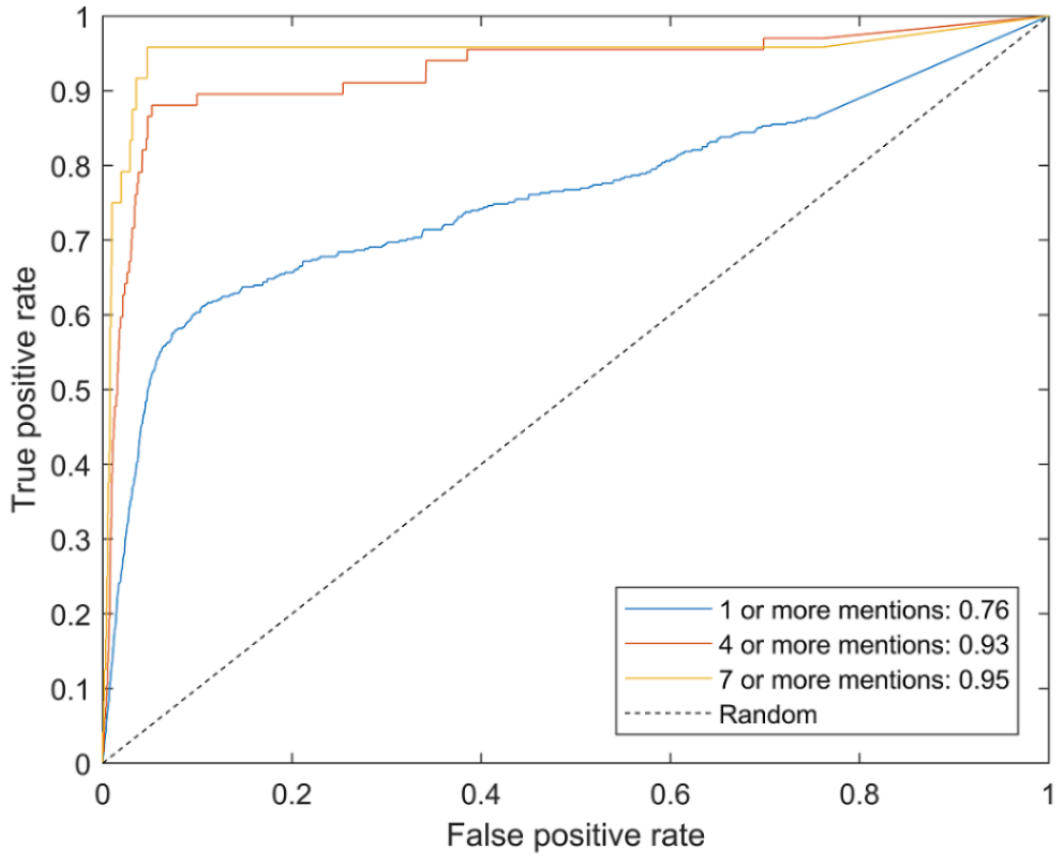
Prospective Cohort

We repeated the same analysis with a prospective cohort comprising users in Dataset 2. Since relatively few people had mentioned that they experienced stroke and were over 60 years of age, our labels were whether or not a user asked about stroke more than a given number of times. This is a weaker proxy than experiential queries, but is nevertheless known to be a good proxy [28]. The control group consisted of all other users.

Our classification model was constructed using data from Dataset 1. Data for the prospective cohort comprised queries until 14 days before a gap followed by the first mention of a stroke.

Figure 3 shows the ROC curve for the resulting experiment, demonstrating the ability to detect people who will, in future, ask multiple times about stroke.

Figure 3. Receiver operating characteristic curve for the prospective cohort; the area under the curve (AUC) values for the different settings are presented in the legend.



Temporal Change Prior to the Indication of Stroke

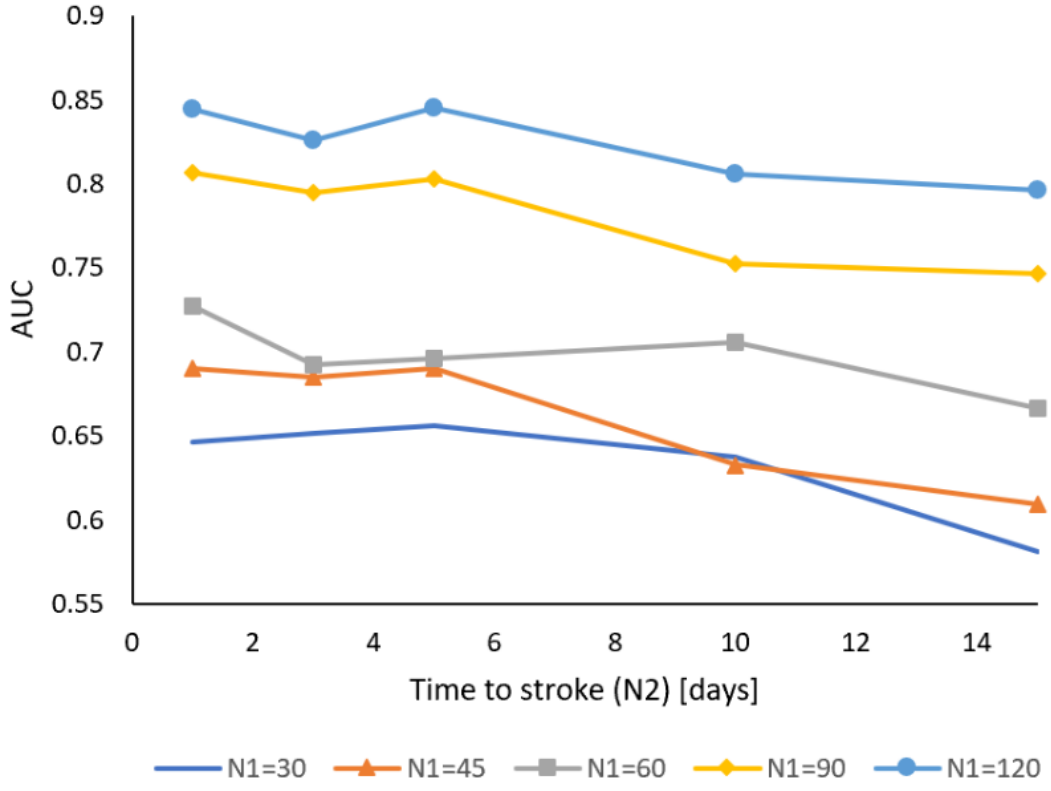
We used the attributes derived for people in the patient cohort from Dataset 1 to examine the ability to localize stroke in time. Toward this end, we considered two time constants, N_1 and N_2, to define the negative examples as data from before N_1 days prior to time zero and positive examples as data from N_2 days prior to time zero until time zero, essentially training a classification model to predict if a user will undergo a stroke in the next N_2 days.

XSL•FO
RenderX

The model was a random forest model with 1000 trees. Ten-fold cross-validation stratified by users was utilized in the estimation of model performance. Figure 4 shows the AUC as a function of N_1 and N_2. Relatively accurate classification could be obtained for the stroke localization task, especially for times close to the event. For example, Figure 4 shows that separation between data coming from 120 or more days before the stroke and data coming from 14 days before the event can be classified with AUC≥0.8. Moreover, two effects are visible. First, as the separation between N_1 and N_2 increases (N_2 is larger), performance is improved. Second, the closer to the date of estimated stroke, the better the prediction. We hypothesize that the signal of stroke becomes stronger closer to the time of the stroke and thus negative training data taken close to the date of stroke confuse the model. Second, by limiting the positive examples to those near the actual stroke date, the data contain stronger indications of impending stroke, making the classification more accurate.

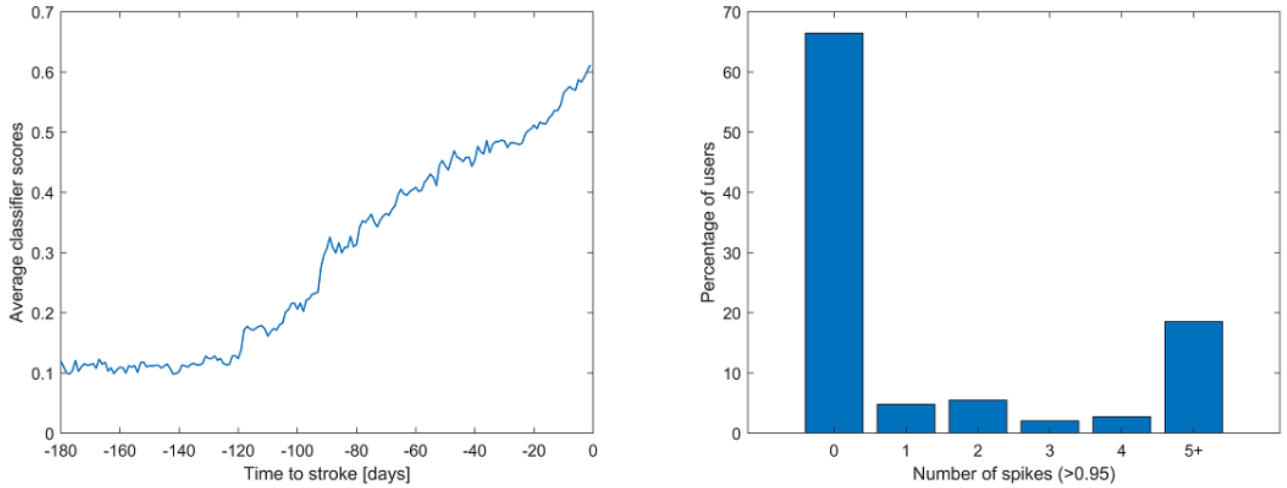
To further investigate this hypothesis, we examined times close to the estimated stroke time with respect to the predicted classifier scores (with N_1=90, N_2=5). Figure 5 shows the average classifier scores (in the range of 0 to 1) as a function of the time relative to the estimated stroke time. The classifier scores began to rise around 120 days before the estimated stroke day. In addition, Figure 5 shows the number of days during the prior 120 days and until the day before the estimated stroke day on which users had a score of over 0.95. A significant percentage of users had 5 or more days on which their scores were extremely high. Taken together, these findings suggest that significant alterations can be recognized prior to the date of estimated stroke.

Figure 4. Area under the receiver operating characteristic curve (AUC) as a function of N_2, the number of days ahead for which the prediction is made. Different curves show different values of N_1, the time prior to stroke that is taken as nonstroke dates.



XSL•FO
RenderX

Figure 5. Average classifier scores as a function of the time to estimated stroke date (left) and the number of times that each user had classifier scores greater than 0.95 (right) in the period between 120 days prior to the estimated stroke date and the day before.



Discussion

Principal Findings

Our findings suggest that among internet users, stroke events are preceded by alterations in communication patterns that have been previously shown to reflect aspects of cognitive function [29]. We used an unusual break in the query stream prior to the first mention of a stroke as the likely time of the clinical cerebrovascular event. Assuming that our presumed time of stroke correctly approximates the time of the real event, the internet-derived signal of an impending stroke tends to gain strength as the time of the stroke approaches. If monitored continuously, not only the signal per se but also its persistence and intensification with time might comprise a useful aid in attempted identification of an impending stroke. We expect that these or similar changes in pattern will be present in subjects at risk for stroke prior to a clinically verified stroke. Clearly, following the present hypothesis-forming analysis, the usefulness of the prediction model described here must be validated, and perhaps improved in a clinical setting among internet users who underwent a documented stroke.

Since stroke is mostly a condition affecting the older population segments, the complexity of multiple confounding morbidities comprised a real challenge to our study. We therefore included four types of control groups: (1) subjects not mentioning any health problems; (2) subjects mentioning acute nonvascular medical conditions (ie, surgery); (3) subjects who mentioned an acute vascular event, acute myocardial infarction with risk factors that are nearly identical to those of stroke; and (4) a variety of chronic conditions, which are relatively common in older subjects and may affect cognition, including hypertension, depression, vitamin B12 deficiency, and hypothyroidism. The model developed in this study performed well against all comparator groups, but showed better distinction from surgery and acute myocardial infarction, with more limited power to separate the stroke signal from that generated by hypertension. However, even in the weakest comparisons, the performance via the ROC analysis was very high, with an AUC in excess of 0.985 for all comparators. To place this performance in the proper context, the revised Framingham Stroke Risk Score, which is based on traditional risk factors and predicts a 10-year risk of stroke, offers an AUC of ~0.71 at best [3].

This study cannot uncover the mechanism underlying the evolution, prior to stroke, of escalating cognitive signals that can be detected via examination of an internet communication pattern. A recognizable stroke could be preceded by showering of microemboli from the arterial tree or the heart; worsening of brain small-vessel disease by uncontrolled hypertension and/or hyperlipidemia; or former smaller, clinically unrecognized microstrokes such as in the case of paroxysmal atrial fibrillation or in situ thrombi with cycles of partial resolution and expansion, all subtly impairing cognitive function prior to a major event. Such putative mechanisms would be hard to verify/exclude in the absence of solid clinical, biochemical, and imaging information.

There appears to be a three-fold advantage to an internet-based app for predicting impending stroke such as that presented here. First, it has the potential of serving as an alarm signal, which becomes gradually more intense as the time of the actual stroke approaches, beginning at 120 days to a few days before the stroke. This offers a time window of opportunity to act swiftly, identify specific underlying conditions that predispose to stroke, and apply preventive measures. This concept potentially generates a concrete and immediate alarm, vastly remote from calculating a general “sometime in the next decade” warning, which is often ignored. Second, if indeed validated in the clinical setting, its predicted overall performance and accuracy are considerably higher than those of solidly established traditional tools. Lastly, this algorithm can be applied in an inexpensive manner on a large scale and ongoing basis, similar to a computer virus detection program that runs in the background, reducing costly screening, morbidity, and mortality [17].

Limitations

This study has important inherent limitations. First, the assumption that subjects who declared that they had a stroke indeed experienced some form of true cerebrovascular event lacks actual clinical documentation. However, we provide indirect evidence that the identification was likely correct, including a period of “internet silence” suggesting

XSL•FO
RenderX

hospitalization; queries for risk factors and relevant drugs, which increased substantially in the stroke cohort after the event; increase in classifier scores leading up to the estimated stroke time; and the attributes used to separate the stroke cohort from the controls. Moreover, many other people who underwent a stroke event did not identify themselves as having experienced it. Those people would not have been included in our cohort and might therefore introduce a bias toward specific populations [30].

Second, since the terms used to trap the occurrence of stroke were selected by lay people, the precise nature of the event is impossible to determine, neither is the ratio among various forms of cerebrovascular events such as true stroke vs transient ischemic attack. The requirement for a minimal period of internet "absenteeism" as an inclusion criterion would likely, but not completely, exclude very minor events. As is the case for the occurrence of stroke itself, the existence of the comparator conditions such as acute myocardial infarction completely depends on wording in queries and lack actual medical evidence.

Third, our analysis only applies to presumed stroke survivors who could return to use the internet. Therefore, it does not apply to individuals who did not survive the stroke or had an unfortunate clinical course that precluded further internet communications after the stroke; thus, our analysis was likely not applied to the most severe forms of stroke. Although it is possible that subjects that suffered fatal or severely disabling stroke would have no prestroke warning fingerprints in their communications, it is equally reasonable to assume that their prestroke signals could be even stronger. Moreover, the actual use of drugs is likely higher than their mention in communications and could also refer to people other than the subjects under study themselves.

We also note that the identification of people in the control condition, although supported by precedent [28], does not preclude people who queried for events that happened in the past, because they were asking about other people or those who were asking out of general interest. However, as the cited literature suggests, they are a reasonable proxy for such a condition cohort.

Conclusions

We have developed an internet-based detection model that retrospectively identified subjects who later developed stroke according to self-reports. The usefulness of this system in real-life patients awaits validation in a clinical setting where this model must be tested against accurate diagnoses, information on stroke risk factors, comorbidities, drugs, and outcome. We put forth the vision, which requires much testing but deserves further pursuit, that consenting internet users at some risk for stroke would request continuous surveillance, much like standard computer antivirus programs, to detect an impending stroke so as to allow its prevention.

Conflicts of Interest

EYT is an employee of Microsoft, owner of Bing. The other authors declare no conflicts of interest.

References

1. Cheng X, Yang Y, Schwebel DC, Liu Z, Li L, Cheng P, et al. Population ageing and mortality during 1990-2017: A global decomposition analysis. PLoS Med 2020 Jun;17(6):e1003138 [FREE Full text] [doi: 10.1371/journal.pmed.1003138] [Medline: 32511229]
2. Reitz C, Bos MJ, Hofman A, Koudstaal PJ, Breteler MM. Prestroke cognitive performance, incident stroke, and risk of dementia: the Rotterdam Study. Stroke 2008 Jan;39(1):36-41. [doi: 10.1161/STROKEAHA.107.490334] [Medline: 18006863]
3. Flueckiger P, Longstreth W, Herrington D, Yeboah J. Revised Framingham Stroke Risk Score, nontraditional risk markers, and incident stroke in a multiethnic cohort. Stroke 2018 Feb;49(2):363-369 [FREE Full text] [doi: 10.1161/STROKEAHA.117.018928] [Medline: 29311270]
4. Egan BM, Zhao Y, Axon RN. US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008. JAMA 2010 May 26;303(20):2043-2050. [doi: 10.1001/jama.2010.650] [Medline: 20501926]
5. McCabe PJ, Barton DL, DeVon HA. Older adults at risk for atrial fibrillation lack knowledge and confidence to seek treatment for signs and symptoms. SAGE Open Nurs 2017;3:1-9 [FREE Full text] [doi: 10.1177/2377960817720324] [Medline: 30637335]
6. Stroke Prevention in Atrial Fibrillation Study Group Investigators. Preliminary report of the Stroke Prevention in Atrial Fibrillation Study. N Engl J Med 1990 Mar 22;322(12):863-868. [doi: 10.1056/NEJM199003223221232] [Medline: 2407959]
7. Staessen JA, Thijisq L, Fagard R, Celis H, Birkenhäger WH, Bulpitt CJ, Systolic Hypertension in Europe (Syst-Eur) Trial Investigators. Effects of immediate versus delayed antihypertensive therapy on outcome in the Systolic Hypertension in Europe Trial. J Hypertens 2004 Apr;22(4):847-857. [doi: 10.1097/00004872-200404000-00029] [Medline: 15126928]
8. Vasiljeva I, Arandjelovic O. Towards sophisticated learning from EHRs: increasing prediction specificity and accuracy using clinically meaningful risk criteria. 2016 Aug Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2016; Orlando, FL p. 2452-2455. [doi: 10.1109/EMBC.2016.7591226]
9. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk:

XSL•FO
RenderX

a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S49-S73. [doi: 10.1161/01.cir.0000437741.48606.98] [Medline: 24222018]
10. Seelye A, Hagler S, Mattek N, Howieson DB, Wild K, Dodge HH, et al. Computer mouse movement patterns: A potential marker of mild cognitive impairment. Alzheimers Dement (Amst) 2015 Dec 01;1(4):472-480 [FREE Full text] [doi: 10.1016/j.dadm.2015.09.006] [Medline: 26878035]
11. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66(Suppl 1):i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]
12. Rodríguez García PL, Rodríguez García D. Letter by Rodríguez-García and Rodríguez-García [corrected] regarding article, "Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association". Stroke 2011 Nov;42(11):e584. [doi: 10.1161/STROKEAHA.111.634279] [Medline: 21940969]
13. Austin J, Hollingshead K, Kaye J. Internet searches and their relationship to cognitive function in older adults: cross-sectional analysis. J Med Internet Res 2017 Sep 06;19(9):e307 [FREE Full text] [doi: 10.2196/jmir.7671] [Medline: 28877864]
14. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006;2006:244-248 [FREE Full text] [Medline: 17238340]
15. 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]
16. Pelleg D, Yom-Tov E, Maarek Y. Can you believe an anonymous contributor? On truthfulness in Yahoo! Answers. 2012 Presented at: International Conference on Privacy, Security, Risk and Trust and 2012 International Confernce on Social Computing; September 3-5, 2012; Amsterdam, Netherlands. [doi: 10.1109/SocialCom-PASSAT.2012.13]
17. Kaye J, Mattek N, Dodge HH, Campbell I, Hayes T, Austin D, et al. Unobtrusive measurement of daily computer use to detect mild cognitive impairment. Alzheimers Dement 2014 Jan;10(1):10-17 [FREE Full text] [doi: 10.1016/j.jalz.2013.01.011] [Medline: 23688576]
18. Youngmann B, Allerhand L, Paltiel O, Yom-Tov E, Arkadir D. A machine learning algorithm successfully screens for Parkinson's in web users. Ann Clin Transl Neurol 2019 Dec;6(12):2503-2509. [doi: 10.1002/acn3.50945] [Medline: 31714022]
19. Soldaini L, Yom-Tov E. Inferring individual attributes from search engine queries and auxiliary information. 2017 Presented at: 26th International Conference on World Wide Web; 2017; Perth, Australia. [doi: 10.1145/3038912.3052629]
20. White RW, Horvitz E. Evaluation of the feasibility of screening patients for early signs of lung carcinoma in web search logs. JAMA Oncol 2017 Mar 01;3(3):398-401. [doi: 10.1001/jamaoncol.2016.4911] [Medline: 27832243]
21. Yom-Tov E. Screening for cancer using a learning internet advertising system. ACM Trans Comput Healthcare 2020 Apr 15;1(2):1-13. [doi: 10.1145/3373720]
22. Hochberg I, Daoud D, Shehadeh N, Yom-Tov E. Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. Acta Diabetol 2019 Oct;56(10):1149-1154. [doi: 10.1007/s00592-019-01350-5] [Medline: 31093762]
23. 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]
24. Guntuku SC, Schwartz HA, Kashyap A, Gaulton JS, Stokes DC, Asch DA, et al. Variability in language used on social media prior to hospital visits. Sci Rep 2020 Mar 12;10(1):4346. [doi: 10.1038/s41598-020-60750-8] [Medline: 32165648]
25. Wang Q, Capistrant BD, Ehntholt A, Glymour MM. Long-term rate of change in memory functioning before and after stroke onset. Stroke 2012 Oct;43(10):2561-2566 [FREE Full text] [doi: 10.1161/STROKEAHA.112.661587] [Medline: 22935399]
26. Zheng F, Yan L, Zhong B, Yang Z, Xie W. Progression of cognitive decline before and after incident stroke. Neurology 2019 Jul 02;93(1):e20-e28. [doi: 10.1212/WNL.0000000000007716] [Medline: 31127071]
27. Thompson M, Duda RO, Hart PE. Pattern Classification and Scene Analysis, 2nd Edition. In: Leonardo. New Yok: Wiley; 1974.
28. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
29. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785. [doi: 10.2196/12785] [Medline: 31471958]
30. Yom-Tov E. Demographic differences in search engine use with implications for cohort selection. Inf Retrieval J 2019 Jan 1;22(6):570-580. [doi: 10.1007/s10791-018-09349-2]

Abbreviations

AUC: area under the receiver operating characteristic curve
ROC: receiver operating characteristic

XSL•FO
RenderX

Edited by G Eysenbach; submitted 11.01.21; peer-reviewed by A Mavragani; comments to author 01.02.21; revised version received 03.02.21; accepted 04.05.21; published 28.05.21.

Please cite as:
Shaklai S, Gilad-Bachrach R, Yom-Tov E, Stern N
Detecting Impending Stroke From Cognitive Traits Evident in Internet Searches: Analysis of Archival Data
J Med Internet Res 2021;23(5):e27084
URL: https://www.jmir.org/2021/5/e27084
doi:10.2196/27084
PMID:34047699

©Sigal Shaklai, Ran Gilad-Bachrach, Elad Yom-Tov, Naftali Stern. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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 Uptake and Use of Telemonitoring in Chronic Care Between 2014 and 2019: Nationwide Survey Among Patients and Health Care Professionals in the Netherlands

Martine W J Huygens[1], MSc, PhD; Helene R Voogdt-Pruis[2], MSc, PhD; Myrah Wouters[3], MA; Maaike M Meurs[1], MSc, PhD; Britt van Lettow[4], MSc, PhD; Conchita Kleijweg[4], MSc; Roland D Friele[1], MSc, PhD

[1]NIVEL Netherlands Institute for Health Services Research, Utrecht, Netherlands

[2]Department of Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[3]National Institute for Public Health and the Environment, Bilthoven, Netherlands

[4]Nictiz, The Hague, Netherlands

Corresponding Author:
Helene R Voogdt-Pruis, MSc, PhD
Department of Global Health
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
PO Box 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 0651964815
Email: h.r.pruis@umcutrecht.nl

Abstract

Background: Telemonitoring could offer solutions to the mounting challenges for health care and could improve patient self-management. Studies have addressed the benefits and challenges of telemonitoring for certain patient groups.

Objective: This paper will examine the nationwide uptake of telemonitoring in chronic care in the Netherlands from 2014 to 2019 by means of an annual representative survey among patients and health care professionals.

Methods: Between 2014 and 2019, approximately 2900 patients with chronic diseases, 700 nurses, and 500 general practitioners (GPs) and medical specialists received a questionnaire. About 30 questions addressed topics about the use of eHealth and experiences with it, including data about telemonitoring.

Results: Between 2014 and 2019, the use of telemonitoring remained stable for all groups except medical specialists. In medical specialist departments, the use of telemonitoring increased from 11.2% (18/161) in 2014 to 19.6% (36/184) in 2019 (χ^2_4=12.3; *P*=.02). In 2019, telemonitoring was used by 5.8% (28/485) of people with chronic disease. This was 18.2% (41/225) in GP organizations and 40.4% (44/109), 38.0% (78/205), and 8.9% (29/325) in the organizations of nurses working in primary, secondary, and elderly care, respectively. Up to 10% of the targeted patient group such as diabetics were regarded by health care professionals as suitable for using telemonitoring. The main benefits mentioned by the patients were “comfort” (421/1043, 40.4%) and “living at home for longer/more comfortably” (334/1047, 31.9%). Health care professionals added “improvement of self-management” (63/176, 35.8% to 57/71, 80.3%), “better understanding of the patient’s condition” (47/176, 26.7% to 42/71, 59.2%), “reduction of workload” (53/134, 39.6% of nurses in elderly care), “better tailoring of care plan to the patient’s situation” (95/225, 42.2% of GPs), and “saves time for patients/caregivers” (61/176, 34.7% of medical specialists). Disadvantages mentioned by professionals were that “it takes time to monitor data” (13/130, 10% to 108/225, 48.0%), “it takes time to follow up alerts” (15/130, 11.5% to 117/225, 52.0%), and “it is difficult to estimate which patients can work with telemonitoring” (22/113, 19.5% to 94/225, 41.8%).

Conclusions: The uptake of telemonitoring in Dutch chronic care remained stable during 2014-2019 but increased among medical specialists. According to both patients and professionals, telemonitoring improves the quality of life and quality of care. Skills for suitably including eligible patients and for allocating the tasks of data monitoring and follow-up care within the team would help to further increase the use of telemonitoring.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e24908) doi:10.2196/24908

KEYWORDS

eHealth; telemonitoring; self-management; telemedicine; telehealth

Introduction

Added Value of Telemonitoring

Telemonitoring could broaden access to health care and offer solutions to the mounting challenges for the health care system such as an ageing population, which is creating a demand for long-term care, rising expectations from patients who are better informed about health issues, and the pressure on national health care budgets due to these demands [1-4]. Telemonitoring uses technology such as videoconferencing, email, remote electronic monitoring equipment, social network apps, and internet portals to allow monitoring and self-monitoring of health data by patients and health-related education and long-distance interventions by health care professionals (HCPs) [5-7]. Several studies have addressed the benefits of telemonitoring, for example, better access to health care and the cost-effective delivery of health care. Telemonitoring could reduce face-to-face consultations and clinic visits. In addition, telemonitoring improves the quality of care and clinical outcomes through continuous and reliable monitoring of data, immediate assessment, triage, and interventions. Telemonitoring could also improve patient empowerment, self-management, and compliance [4,8-16].

Implementation of Telemonitoring

Implementation and actual use of telemonitoring in daily practice require well-thought-out action plans, for example, for selecting appropriate interventions or for tailoring the design to the needs of the user group. In addition, telemonitoring should be seamlessly integrated into the health care processes and should avoid disrupting the HCPs' existing workflow. It was also recommended that telemonitoring should be part of "blended care" as both patients and HCPs prefer face-to-face encounters [17,18]. Implementation programs for telemonitoring should tackle barriers to actual use. Some studies found barriers that are related to "users" of telemonitoring such as lack of digital skills, resistance to change, and lack of direct personal benefit. Mentioned examples of barriers that were related to the context were lack of privacy (or fear of it), security, patient safety, a properly working internet connection, proper technological infrastructure, regulations, funding, and task allocation [4,19].

Telemonitoring in the Netherlands

In the Netherlands, an enabling factor related to the context of telemonitoring is the availability of 4G mobile networks and high-speed broadband Internet access, even in rural areas [20]. Currently, 90% of Dutch people use the internet daily. In particular, over the last 5 years, older people, the less well-educated, those born outside the country, and low-income households have caught up [21]. In terms of policy, several national documents, studies, and guidelines for eHealth have been developed during the last 10 years addressing privacy and patient safety in the use of eHealth [22,23]. In 2012, the Dutch National Implementation Agenda for eHealth was launched, followed by the eHealth Governance Covenant 2014 - 2019 [24] and a framework on the use of eHealth by HCPs [23]. As regards funding, telemonitoring is covered by the Dutch health insurance system; in particular, the funding of follow-up consultations improved in 2019. Still, HCPs do not have the funds to upscale telemonitoring [25].

Annual Nationwide Representative Study Among Health Care Professionals

Since 2013, the nationwide uptake of eHealth in general by patients and HCPs has been investigated annually and reported in what is known as the "eHealth-monitor." This investigation was commissioned by the Dutch Ministry of Health, Welfare and Sport. The aim of the annual eHealth-monitor was to investigate the implementation of eHealth and to boost its implementation in subsequent health care policy. Every year, about 30 questions addressed topics on the use of eHealth and experiences with it, including telemonitoring. The findings from the perspectives of nurses, general practitioners (GPs), medical specialists, and patients with a chronic disease increased the understanding of the implementation of telemonitoring and the uptake of telemonitoring in daily practice. To our knowledge, our study is the first scientific paper on the nationwide uptake of telemonitoring for all patient groups in chronic care over a long period of time. We analyzed data on the actual use of telemonitoring by patients and HCPs in daily chronic care. In addition, opinions on and experiences with telemonitoring were analyzed.

Methods

Study Design

Since 2013, data for the eHealth-monitor has been collected annually using various nationwide panels. Written and online questionnaires on eHealth were sent to GPs, medical specialists, nurses (practice nurses and practice assistants working in elderly care, GP care, and hospital care), and people with chronic diseases. All participants were approached in March. Nonresponders initially received 1 written or 2 online reminders. For this study, data from respondents on questions about telemonitoring between 2014 and 2019 were used.

Study Population

People with chronic diseases may be included in the representative National Panel of people with Chronic illness or Disability (NPCD) [26]. Inclusion criteria for the NPCD are age 15 or older, diagnosed with a somatic chronic disease, aware of the diagnosis, having a life expectancy of more than 6 months, mentally capable of participating, and not permanently institutionalized. Every year, 500 new panel members are selected to replace panel members who have withdrawn or who have participated for 4 years. The questions on telemonitoring were posed in 2015, 2017, and 2019.

XSL•FO
RenderX

Nurses are participants in a representative nursing staff panel. The nursing staff panel consists of a nationwide group of nursing staff members (nurses, caregivers, and practice assistants) in various health care settings who deliver direct patient care. The recruitment of members of the nursing staff panel takes place through a random sample of two pension funds. Together, these pension funds register all employees in the Dutch health care sector. Nursing staff were asked to participate in health care research for various purposes. People who agreed and who delivered direct nursing care to patients could join the nursing staff panel. The questions on telemonitoring were asked annually, from 2014 until 2019. For this study, the data of nurses working in primary care, secondary care, and elderly care were used. For 2014 until 2016, the data of nurses working in the curative sectors (practice nurses and nurses working in hospital care) were taken together. From 2017 onwards, these sectors were split into two samples.

General practitioners and medical specialists are participants in a representative doctor's panel. Included are all registered GPs and medical specialists of the Royal Dutch Medical Association. Inclusion criteria for participating in the eHealth-monitor were practicing in the past year and being involved in the diagnosis or treatment of patients. From these doctors, certain specializations were excluded: public and occupational health, forensic medicine, addiction medicine, and psychiatry. The questions on telemonitoring were asked annually, from 2014 until 2019.

Questionnaires

All participants were asked about their use of telemonitoring and experience with its advantages and disadvantages in the previous 12 months. In addition, HCPs were asked for which portion of patients telemonitoring was used and for which patient groups relevant (Textbox 1).

XSL•FO
RenderX

Textbox 1. Questionnaires.

Telemonitoring: Remote monitoring of a patient, in which they measure their own health values (for example, blood pressure, blood sugar level) using a meter, sensor, or other device in the home situation and in which they could also respond to some questions. The HCP receives these data digitally.

People with chronic diseases who measure health values themselves

- Which of the following statements apply to you? (multiple answers possible)
- I electronically submit my self-measured health values to my health care provider (eg, by email or automatically via computer or mobile app) (2015, 2017, 2019)
- My health care provider can see my health data on a website or in a mobile app (2015, 2017, 2019)
- My health care provider looks at my self-measured health data before or during a consultation and discusses it with me (2015, 2017, 2019)
- My health care provider keeps an eye on my self-measured health data remotely and contacts me if anything is wrong (2015, 2017, 2019)
- Can you say how desirable or necessary telemonitoring is for you? (2017)
- Could you please answer the following statements? (2019) I notice or think that telemonitoring...
  - ...makes it easier for me to live at home longer and/or more easily
  - ...takes a lot of effort for me
  - ...improves my care
  - ...makes me very tense

Nurses

- Has telemonitoring been used in your organization in the past year? (2014-2019)
- If so, what proportion of your clients/patients use telemonitoring (estimated)? (2016, 2017, 2019)
- If so, what proportion of your clients do you think telemonitoring makes sense for (estimated)? (2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

General practitioners

- Could you state whether telemonitoring is applied to the following patient groups (in your practice)? If telemonitoring is not applied, could you please state whether there are plans to start within a year and whether you would like to do so? (2014-2017)
- In your practice, some of your patients with diabetes are currently being monitored by telemonitoring. Could you please estimate the proportion of your diabetes patients for whom you think telemonitoring is sensible? What proportion of your diabetes patients do use it? (2015-2017)
- Is telemonitoring of patients relevant to you or your practice, and is it used within your practice for some or all of the patients? (2019)
- Could you estimate the size of the group of patients for whom you think telemonitoring is sensible and for whom you or other HCPs actually use it in your practice? (2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

Medical specialists

- Is telemonitoring of patients relevant to your medical specialty, and is it applied by your department for some or all patients? (2014-2017, 2019)
- For which group of your patients is telemonitoring relevant? (2015-2017)
- Could you estimate the size of the group of patients for whom you think telemonitoring is sensible and for whom you or other HCPs actually use it at your department? (2015-2017, 2019)
- What advantages and disadvantages do you experience with telemonitoring or expect from it? (multiple answers possible) (2019)

Data Analysis

Descriptive analyses were conducted to study the use and experiences of telemonitoring. Data from the questionnaires among people with a chronic disease and nursing staff were analyzed using Stata, version 15.0 (StataCorp). The results from the questionnaires among GPs and medical specialists were analyzed using SPSS, version 25.0 (IBM Corp).

For questions asked to people with chronic diseases, the descriptive analyses were weighted for age and gender in such a way that it resembled the distribution of age and gender within the Dutch population from age 18 years, based on data from Statistics Netherlands. We applied a weighting factor ranging from 0.65 to 2.28. The samples for nurses and GPs are fairly representative of the Dutch population of nurses and GPs regarding gender, but the response is not representative for age: GPs younger than 35 years and GPs and nurses 50 years and

XSL•FO
RenderX

older responded more often. Nevertheless, we did not use a weight factor to correct for this because applying the weight factor did not affect the results. For questions asked to medical specialists, the descriptive analyses were weighted for type of specialty. We applied a weighting factor ranging from 0.5 to 1.7.

Results

Participants

Over the years, data were used from 485-633 people with a chronic disease, 322-607 nurses working in elderly care, 220-367 nurses working in the curative disciplines (primary and secondary care), 225-396 GPs, and 184-386 medical specialists who answered questions about telemonitoring (Table 1). The mean age of people with chronic disease was 64.5 to 66.4 years. The mean age of nurses varied from 46.7 to 52.2 years. The mean ages of GPs and medical specialists ranged from 50.0 to 52.6 years and 49.4 to 53.9 years, respectively. Approximately half of the people with chronic disease and the doctors were men; among nurses, only 3.5% (21/607) to 16.1% (25/155) were men.

XSL•FO
RenderX

Table 1. Study population per year.[a]

Characteristic	2014	2015	2016	2017	2018	2019
People with chronic disease						
Responses, n	__[b]	1448	—	1357	—	1292
Telemonitoring responses, n[c]	—	604	—	633	—	485
Male, n (%)	—	315 (52.2)	—	304 (48.0)	—	234 (48.2)
Age (years), mean (SD)	—	64.9 (12.7)	—	66.4 (12.1)	—	64.5 (11.9)
Low education, n (%)	—	133 (22.0)	—	156 (24.6)	—	95 (19.6)
Nurses in elderly care						
Telemonitoring responses, n	408	607	433	341	322	325
Male, n (%)	18 (4.4)	21 (3.5)	20 (4.6)	15 (4.4)	29 (9.0)	19 (5.8)
Age (years), mean (SD)	50.8 (8.9)	48.9 (10.8)	49.7 (10.3)	52.2 (9.5)	50.1 (10.8)	50.9 (11.1)
Nurses in curative sectors						
Telemonitoring responses, n	262	316	220	—	—	—
Male, n (%)	26 (9.9)	36 (11.4)	25 (11.4)	—	—	—
Age (years), mean (SD)	48.8 (9.6)	46.7 (11.4)	48.4 (10.7)	—	—	—
Practice nurses						
Telemonitoring responses, n	—	—	—	212	124	109
Male, n (%)	—	—	—	12 (5.7)	8 (6.5)	8 (7.3)
Age (years), mean (SD)	—	—	—	50.8 (8.6)	50.4 (9.0)	52.0 (9.1)
Nurses in hospital care						
Telemonitoring responses, n	—	—	—	155	211	205
Male, n (%)	—	—	—	25 (16.1)	29 (13.7)	27 (13.2)
Age (years), mean (SD)	—	—	—	49.4 (11.0)	50.0 (11.2)	48.2 (12.4)
General practitioners						
Telemonitoring responses, n	—	396	316	290	—	225
Male, n (%)	—	194 (49.0)	162 (51.3)	133 (45.9)	—	106 (47.1)
Age (years), mean (SD)	—	50.0 (10.2)	51.1 (9.9)	51.5 (9.7)	—	52.6 (9.1)
Medical specialists						
Telemonitoring responses, n[d]	—	386	274	253	—	184
Male, n (%)	—	203 (52.6)	151 (55.1)	153 (60.5)	—	118 (64.1)
Age (years), mean (SD)	—	49.8 (11.3)	49.4 (11.1)	51.9 (11.3)	—	53.9 (9.7)

[a]Among general practitioners and medical specialists, the questions in 2015, 2016, and 2017 were different from those in 2019.

[b]Not available.

[c]Question was only asked to persons who measured health outcomes by themselves.

[d]Weighted by type of specialty.

Actual Use of Telemonitoring

In 2019, 5.8% (28/485) of people with a chronic disease stated that their HCP monitors the health values remotely and contacts them if anything looks wrong (Table 2). In 2015, this figure was 3.3% (20/604). Most of them had been diagnosed with cardiovascular disease or diabetes. In 2019, 38.0% (78/205) of nurses working in secondary care and 40.4% (44/109) in primary care stated that their organization uses telemonitoring. In 2014, this percentage was 34.0% (89/262) for nurses working in the curative disciplines (both hospital and general practice nurses). Among nurses working in elderly care this was 8.9% (29/325) in 2019 and 11.3% (46/408) in 2014. In addition, 18.2% (41/225)of GPs and 19.6% (36/184)of medical specialists stated that telemonitoring was used in their organizations in 2019. These percentages were 17.0% and 11.2%, respectively, in 2014. Among medical specialists, this number has grown significantly over the years (χ^2_4=12.3, P=.02). Up to 10% of the targeted

XSL•FO
RenderX

patient group, including people with diabetes, is reckoned to be suitable for using telemonitoring.

Table 2. Proportions of patients and health care professionals (HCPs) using telemonitoring, from 2014 to 2019.

Population using telemonitoring	2014	2015	2016	2017	2018	2019
Patients with a chronic disease, n (%)[a]	—[b]	20 (3.3)	—	37 (5.8)	—	28 (5.8)
Elderly care						
Nurses, n (%)[c]	46 (11.3)	70 (11.5)	—	—	—	29 (8.9)
Patients using telemonitoring, n (mode %)[d]	—	—	27 (1-10)	25 (1-10)	—	17 (1-10)
Primary/secondary care						
Nurses, n (%)	89 (34.0)	126 (39.9)	84 (38.2)	—	—	—
Patients using telemonitoring, n (mode %)	—	—	18 (1-10)	—	—	—
Primary care						
Nurses, n (%)	—	—	—	85 (40.1)	58 (46.8)	44 (40.4)
Patients using telemonitoring, n (mode %)	—	—	—	61 (1-10)	—	25 (1-10)
Secondary care						
Nurses, n (%)	—	—	—	50 (32.3)	80 (37.9)	78 (38.0)
Patients using telemonitoring, n (mode %)	—	—	—	14 (1-10)	—	22 (1-10)
General practitioners, n (%)[c]	29 (17.0)	49 (12.4)	41 (13.0)	26 (9.0)	—	41 (18.2)
Group using telemonitoring, n (mode group)	25 (Diabetes)	43 (Diabetes)	37 (Diabetes)	21 (Diabetes)	—	—
Relevant for diabetes, n (mode %)[d]	—	26 (1-10)	20 (1-10)	10 (1-10)	—	—
Medical specialists, n (%)[c]	18 (11.2)	41 (10.6)	26 (9.5)	29 (11.5)	—	36 (19.6)
Relevant for patients, n (mode %)[d]	—	24 (1-10)	19 (1-10)	—	—	—
Relevant for patient group, n (mode group)[e]	—	13 (Diabetes)	9 (Diabetes)	11 (Diabetes)	—	—

[a]The proportion of responding patients stating they are monitored on self-reported health measures by HCP (remotely).

[b]Not available.

[c]The proportion of responding HCPs stating that telemonitoring is relevant or is used at their department, practice, or organization. Difference between years (χ^2_4=12.3; *P*=.02).

[d]Proportion of patients using telemonitoring according to the HCPs; possible answers: none, up to 10%, up to 20%, up to 50%, up to 100%, I don't know.

[e]The most often mentioned were these groups: diabetes, heart failure, chronic obstructive pulmonary disease, and asthma.

Advantages and Disadvantages of Telemonitoring

Of patients with a chronic disease, 40.4% (421/1043) agreed or totally agreed with the statement "telemonitoring improves my comfort" (353/1043, 33.8% answered "don't know"). In addition, 31.9% (334/1047) agreed or totally agreed with "telemonitoring lets me stay at home longer and/or live more comfortably" (405/1050, 38.7% answered "don't know") (Table 3).

On the other hand, 7.7% agreed or totally agreed with "telemonitoring takes me a lot of effort" (36.4% answered "don't know"). Among HCPs, the most widely experienced or expected advantage of telemonitoring mentioned was "telemonitoring improves patients' self-management" (57/71, 80.3% of nurses working in primary care; 71/124, 57.3% of nurses working in secondary care; 84/134, 62.7% of nurses working in elderly care; 135/225, 60% of GPs; 63/176, 35.8% of medical specialists) (Table 4). In addition, 59.2% (42/71) of nurses working in primary care and 46.8% (58/124) of nurses working in secondary care stated that their experience or expectation of telemonitoring was that they would get a better understanding of the health condition of the patient. Of nurses working in elderly care, 39.6% (53/134) experienced or expected telemonitoring to reduce the workload. Moreover, 42.2% (95/225) of GPs said that telemonitoring let them tailor the care plan to the situation of their patients better. Of medical specialists, 34.7% (61/176) expected or had observed that telemonitoring gave them spare time for patients and caregivers/relatives.

XSL•FO
RenderX

Table 3. Opinions of telemonitoring from patients with chronic diseases (n=1023-1050), 2019.

I notice or think that telemonitoring:	Agreement with statement, n (%)					
	Totally disagree	Disagree	Neither agree nor disagree	Agree	Totally agree	I don't know
Lets me live at home longer and/or more easily (n=1047)	47 (4.5)	62 (5.9)	199 (19.0)	235 (22.4)	99 (9.5)	405 (38.7)
Improves my comfort (n=1043)	42 (4.0)	50 (4.8)	177 (17.0)	331 (31.7)	90 (8.6)	353 (33.8)
Takes me a lot of effort (n=1023)	83 (8.1)	249 (24.3)	246 (24.0)	61 (6.0)	12 (1.2)	372 (36.4)
Improves my care (n=1035)	41 (4.0)	66 (6.4)	234 (22.6)	241 (23.3)	82 (7.9)	371 (35.8)
Makes me very tense (n=1027)	104 (10.1)	220 (21.4)	238 (23.2)	77 (7.5)	17 (1.7)	371 (36.1)

Table 4. Experienced or expected advantages and disadvantages of telemonitoring assessed by health care providers, 2019.

Characteristic	Nurses, primary care	Nurses, secondary care	Nurses, elderly care	General practitioners	Medical specialists
Advantages respondents, n	71	124	134	225	176
Advantages, n (%)[a]					
It reduces the workload	10 (14.1)	23 (18.5)	53 (39.6)	16 (7.1)	14 (8.0)
It reduces the workload of my assistants	—[b]	—	—	56 (24.9)	18 (10.2)
It saves time for patients and or caregivers/relatives	32 (45.1)	58 (46.8)	45 (33.6)	86 (38.2)	61 (34.7)
It improves the quality of care in my organization	34 (47.9)	50 (40.3)	38 (28.4)	77 (34.2)	44 (25.0)
It improves the self-management of the patient	57 (80.3)	71 (57.3)	84 (62.7)	135 (60.0)	63 (35.8)
I have a better understanding of my clients' health condition	42 (59.2)	58 (46.8)	43 (32.1)	77 (34.2)	47 (26.7)
It lets me tailor the care plan better to my patients' situation	38 (53.5)	47 (37.9)	32 (23.9)	95 (42.2)	45 (25.6)
It lets patients ask for help in time	20 (28.2)	52 (41.9)	38 (28.4)	67 (29.8)	32 (18.2)
Other advantages	—	7 (5.6)	3 (2.2)	9 (4.0)	5 (2.8)
I do not expect or experience any advantages	—	9 (7.3)	23 (17.2)	32 (14.2)	48 (27.3)
Disadvantages respondents, n	68	113	130	225	181
Disadvantages, n (%)[a]					
It takes me a lot of time to monitor/check health values	32 (47.1)	21 (18.6)	13 (10.0)	108 (48.0)	56 (30.9)
It takes me a lot of time to follow up notifications	29 (42.6)	19 (16.8)	15 (11.5)	117 (52.0)	51 (28.2)
It takes a lot of time for my assistants	—	—	—	99 (44.0)	45 (24.9)
It ensures that patients and/or relatives contact me more often	20 (29.4)	22 (19.5)	21 (16.2)	83 (36.9)	40 (22.1)
It worries patients and/or relatives	10 (14.7)	20 (17.7)	23 (17.7)	81 (36.0)	33 (18.2)
I find it difficult to work with it	—	—	—	16 (7.1)	7 (3.9)
The system provides unreliable data	—	—	—	52 (23.1)	33 (18.2)
The application is not secure	—	—	—	20 (8.9)	7 (3.9)
I find it difficult to estimate which patients can work with it	15 (22.1)	22 (19.5)	37 (28.5)	94 (41.8)	52 (28.7)
Other disadvantages	8 (11.8)	7 (6.2)	8 (6.2)	18 (8.0)	20 (11.0)
I do not expect or experience any disadvantages	13 (19.1)	49 (43.4)	56 (43.1)	13 (5.8)	33 (18.2)

[a]Multiple answers possible.

[b]Not available.

A total of 43.4% (49/113) of nurses working in secondary care and 43.1% (56/130) of nurses working in elderly care did not expect or experience any disadvantages of telemonitoring. In contrast, 47.1% (32/68) of nurses working in primary care

XSL•FO
RenderX

expected or noted that telemonitoring takes a lot of time to monitor and check health values. Of GPs and medical specialists, 48.0% (108/225) and 30.9% (56/181), respectively, expected or experienced this. In addition, 42.6% (29/68) of nurses working in primary care, 52.0% (117/225) of GPs, and 28.2% (51/181) of medical specialists expected or experienced that using telemonitoring would take up a lot of time following up on alerts. Moreover, 19.5% (22/113) of nurses working in secondary care, 28.5% (37/130) of the nurses working in elderly care, and 28.7% (52/181) of medical specialists expected or experienced difficulties in estimating which patients could handle telemonitoring.

Discussion

Principal Findings

Our study adds new insights to current scientific studies of telemonitoring, as it investigated the actual nationwide uptake of telemonitoring for all patient groups in chronic care over a long period of time (before the COVID-19 pandemic). Findings from the perspectives of nurses, GPs, medical specialists, and patients with chronic diseases can assist the implementation of telemonitoring and the uptake of telemonitoring in daily practice. The current COVID-19 pandemic has called for rapid implementation of telemonitoring for acute and subacute diseases in 2020. Future editions of the Dutch eHealth-monitor might present the impact of COVID-19 on a rising uptake of telemonitoring [27] and on new opinions and experiences with telemonitoring. Strengths of our study include the large sample size, the external validity and reliability of the data, and the representativeness of the various groups of participants (people with chronic diseases, nurses, GPs, and medical specialists). Nevertheless, there are also some limitations. Due to the specific factors in the Netherlands that boosted the implementation of telemonitoring, our study results can only partly be extrapolated to other countries. In addition, the use of telemonitoring was only investigated by asking potential users in health care; we did not investigate data from other resources such as health care insurers or telehealth companies. As well, our quantitative approach is best suited to answering "what," "when," and "who" questions and less well-suited to "how" and "why" questions. The opinions and experiences with telemonitoring that we investigated therefore do not fully explain the factors concerning the implementation and uptake of telemonitoring. Even so, our results concerning the experiences of HCPs are underlined by qualitative studies [28-30]. Other benefits and barriers found are "an increased feeling of safety" and "insufficient familiarity with the technology" [28-30]. In addition, HCPs need to add telemonitoring to the health care process, with a precise description of the target group, task allocation for data monitoring, and support for patients from within the team [31,32].

Conclusion

The uptake of telemonitoring in Dutch chronic care remained stable during 2014-2019 but increased among medical specialists. According to both patients and professionals, telemonitoring improves the quality of life and quality of care. Skills for appropriately including eligible patients and allocating the tasks of data monitoring and follow-up care within the team would help to further increase the use of telemonitoring.

Acknowledgments

The project was funded by the Dutch Ministry of Health, Welfare and Sport. The Dutch Ministry of Health, Welfare and Sport had no role in the collection, analysis, or interpretation of data or in writing the manuscript. The authors would like to acknowledge all other people who were involved in data collection during 2013-2019; in particular, Johan Krijgsman, Ilse Swinkels, Judith de Jong, Janke de Groot, Anne Brabers, Lies van Gennip, Karlijn de Bruin, and Annemarie Lamain. In addition, the authors would like to thank the members of the advisory committee in 2019: Guus Schrijvers, Christine Aberson-Blok, Arina Burghouts, Jeroen Crasborn, Henny Mulders, Henk Herman Nap, Sjaak Nouwt, Renée Verwey, Karin Zwager-Ankone, and Nienke Zwennes.

Authors' Contributions

MWJH and HRVP were the main contributors in data analysis and in writing the manuscript. MW, MWJH, BvL, CK, and RDF were involved in developing the questionnaires. All authors were involved in the process of data management. MWJH, HRVP, MMM, MW, and BvL were involved in the data analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
2. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood) 2010 Aug;29(8):1489-1495. [doi: 10.1377/hlthaff.2009.0888] [Medline: 20679652]
3. European Commission, Liikanen E. The Importance of eHealth in Europe. 2003. URL: https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_03_608 [accessed 2020-01-03]

XSL•FO
RenderX

4. World Health Organization. From innovation to implementation: eHealth in the WHO European Region. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf [accessed 2020-01-07]
5. Telehealth programs. 2018. URL: https://www.hrsa.gov/rural-health/telehealth/index.html [accessed 2019-11-01]
6. Larsen SB, Sørensen NS, Petersen MG, Kjeldsen GF. Towards a shared service centre for telemedicine: Telemedicine in Denmark, and a possible way forward. Health Informatics J 2016 Dec;22(4):815-827 [FREE Full text] [doi: 10.1177/1460458215592042] [Medline: 26261216]
7. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
8. Lopez-Villegas A, Catalan-Matamoros D, Robles-Musso E, Bautista-Mesa R, Peiro S. Cost-utility analysis on telemonitoring of users with pacemakers: The PONIENTE study. J Telemed Telecare 2019 May;25(4):204-212. [doi: 10.1177/1357633X18767184] [Medline: 29621908]
9. Boyne JJ, Vrijhoef HJ, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels AP, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
10. Sul A, Lyu D, Park D. Effectiveness of telemonitoring versus usual care for chronic obstructive pulmonary disease: A systematic review and meta-analysis. J Telemed Telecare 2020 May;26(4):189-199. [doi: 10.1177/1357633X18811757] [Medline: 30541375]
11. Qian W, Lam TT, Lam HHW, Li C, Cheung YT. Telehealth Interventions for Improving Self-Management in Patients With Hemophilia: Scoping Review of Clinical Studies. J Med Internet Res 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: 10.2196/12340] [Medline: 31293241]
12. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
13. 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]
14. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. Int J Med Inform 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004] [Medline: 29029687]
15. Mathavakkannan S, Bowser M, Doyle S, Hoare A, Rehsi G. Promoting patient empowerment and sustainability in kidney care using telemedicine. Int J Integr Care 2014 Nov 01;14(8):2. [doi: 10.5334/ijic.1816]
16. van der Burg JM, Aziz NA, Kaptein MC, Breteler MJ, Janssen JH, van Vliet L, et al. Long-term effects of telemonitoring on healthcare usage in patients with heart failure or COPD. Clinical eHealth 2020;3:40-48. [doi: 10.1016/j.ceh.2020.05.001]
17. Ziefle M, Klack L, Wilkowska W, Holzinger A. Acceptance of telemedical treatments – A medical professional point of view. 2013 Presented at: International conference on human interface and the management of information; 2013; Berlin, Germany p. 21-26. [doi: 10.1007/978-3-642-39215-3_39]
18. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
19. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
20. Eurostat. Digital economy and society statistics - households and individuals. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals [accessed 2020-01-07]
21. Statistics Netherlands (CBS), Eurostat. ICT use by households and individuals. URL: https://opendata.cbs.nl/statline/#/CBS/en/?fromstatweb [accessed 2019-11-01]
22. National Institute for Public Health and Environment (RIVM), Ossebaard HC, de Bruijn ACP, van Gemert-Pijnen JEWC, Geertsma RE. Risk related to the use of ehealth technologies; an explorative study. 2012. URL: https://www.rivm.nl/bibliotheek/rapporten/360127001.pdf [accessed 2020-01-07]
23. Inspectorate on health and youth care (IGJ). Toetsingskader IGJ; inzet van e-health door zorgaanbieders. Report. 2018. URL: https://www.igj.nl/binaries/igj/documenten/toetsingskaders/2019/10/18/toetsingskader-inzet-van-e-health-door-zorgaanbieders/Toetsingskader_IGJ_Inzet+e-health+door+zorgaanbieders.pdf [accessed 2020-01-07]
24. Nivel, Kroneman M, Boerma W, van den Berg M, Groenewegen P, de Jong J, et al. Health systems in transition. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0016/314404/HIT_Netherlands.pdf [accessed 2020-01-07]
25. NIVEL, Meurs M, Huygens M, Keuper J, de Groot J, Friele R. Slimme zorg thuis bij COPD en hartfalen Ervaringen en ambities van vijf voorloper initiatieven In Dutch. 2019. URL: https://nivel.nl/sites/default/files/bestanden/1003591.pdf [accessed 2020-01-07]
26. Rijken M, Bensing J. The National Panel of Patients with Chronic Diseases: An infrastructure for patient-centered research. TSG Tijdschrift Voor Gezondheidswetenschappen 2000;78:5.

XSL•FO
RenderX

27. Lai SHS, Tang CQY. Telemedicine and COVID-19: beyond just virtual consultations - the Singapore experience. Bone Jt Open 2020 Jun;1(6):203-204 [FREE Full text] [doi: 10.1302/2633-1462.16.BJO-2020-0042.R1] [Medline: 33225290]
28. Van der Zweth J, Askari M, Brinkhuis M, van Lettow B. The Added Value of Telehomecare for Heart Patients, Healthcare Professionals, and Healthcare Organizations. 2017. URL: https://dspace.library.uu.nl/bitstream/handle/1874/354694/van%20der%20Zweth%20-%205675383%20-%20Thesis.pdf [accessed 2020-01-07]
29. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
30. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. A qualitative study of chronic obstructive pulmonary disease patient perceptions of the barriers and facilitators to adopting digital health technology. Digit Health 2019;5:2055207619871729 [FREE Full text] [doi: 10.1177/2055207619871729] [Medline: 31489206]
31. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. Int J Chron Obstruct Pulmon Dis 2019;14:1681-1690 [FREE Full text] [doi: 10.2147/COPD.S207187] [Medline: 31440044]
32. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]

Abbreviations

GP: general practitioner
HCP: health care professional
NPCD: National Panel of people with Chronic illness or Disability

Edited by R Kukafka; submitted 09.10.20; peer-reviewed by K Fitzner, O Blanson Henkemans; comments to author 23.11.20; revised version received 27.11.20; accepted 21.03.21; published 03.05.21.

Please cite as:
Huygens MWJ, Voogdt-Pruis HR, Wouters M, Meurs MM, van Lettow B, Kleijweg C, Friele RD
The Uptake and Use of Telemonitoring in Chronic Care Between 2014 and 2019: Nationwide Survey Among Patients and Health Care Professionals in the Netherlands
J Med Internet Res 2021;23(5):e24908
URL: https://www.jmir.org/2021/5/e24908
doi:10.2196/24908
PMID:33938808

©Martine W J Huygens, Helene R Voogdt-Pruis, Myrah Wouters, Maaike M Meurs, Britt van Lettow, Conchita Kleijweg, Roland D Friele. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2021. 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 Costs and Cardiovascular Benefits in Patients With Peripheral Artery Disease From a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study

Jen-Kuang Lee[1,2,3,4,5], PhD, MD; Chi-Sheng Hung[1,3,4,5], PhD, MD; Ching-Chang Huang[1,3,4,5], MD; Ying-Hsien Chen[1,3,4,5], MD; Hui-Wen Wu[6], BN; Pao-Yu Chuang[6], MSN; Jiun-Yu Yu[7], PhD; Yi-Lwun Ho[1,3,4,5], PhD, MD

[1]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[2]Department of Laboratory Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

[3]Telehealth Center, National Taiwan University Hospital, Taipei, Taiwan

[4]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan

[5]Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

[6]Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan

[7]Department of Business Administration, College of Management, National Taiwan University Hospital, Taipei, Taiwan

Corresponding Author:
Yi-Lwun Ho, PhD, MD
Division of Cardiology
Department of Internal Medicine
National Taiwan University Hospital
7 South Chung Shan Road
Taipei,
Taiwan
Phone: 886 2 23123456 ext 62152
Email: ylho@ntu.edu.tw

Abstract

Background: Patients with peripheral artery disease (PAD) are at high risk for major cardiovascular events, including myocardial infarction, stroke, and hospitalization for heart failure. We have previously shown the clinical efficacy of a fourth-generation synchronous telehealth program for some patients, but the costs and cardiovascular benefits of the program for PAD patients remain unknown.

Objective: The telehealth program is now widely used by higher-risk cardiovascular patients to prevent further cardiovascular events. This study investigated whether patients with PAD would also have better cardiovascular outcomes after participating in the fourth-generation synchronous telehealth program.

Methods: This was a retrospective cohort study. We screened 5062 patients with cardiovascular diseases who were treated at National Taiwan University Hospital and then enrolled 391 patients with a diagnosis of PAD. Of these patients, 162 took part in the telehealth program, while 229 did not and thus served as control patients. Inverse probability of treatment weighting (IPTW) based on the propensity score was used to mitigate possible selection bias. Follow-up outcomes included heart failure hospitalization, acute coronary syndrome, stroke, and all-cause readmission during the 1-year follow-up period and through the last follow-up.

Results: The mean follow-up duration was 3.1 (SD 1.8) years for the patients who participated in the telehealth program and 3.2 (SD 1.8) for the control group. The telehealth program patients exhibited lower risk of ischemic stroke than did the control group in the first year after IPTW (0.9% vs 3.5%; hazard ratio [HR] 0.24; 95% CI 0.07-0.80). The 1-year composite endpoint of vascular accident, including acute coronary syndrome and stroke, was also significantly lower in the telehealth program group after IPTW (2.4% vs 5.2%; HR 0.46; 95% CI 0.21-0.997). At the end of the follow-up, the telehealth program group continued to exhibit a significantly lower rate of ischemic stroke than did the control group after IPTW (0.9% vs 3.5%; HR 0.52, 95% CI 0.28-0.93). Furthermore, the medical costs of the telehealth program patients were not higher than those of the control group, whether in terms of outpatient, emergency department, hospitalization, or total costs.

XSL•FO
RenderX

Conclusions: The PAD patients who participated in the fourth-generation synchronous telehealth program exhibited lower risk of ischemic stroke events over both mid- and long-term follow-up periods. However, larger-scale and prospective randomized clinical trials are needed to confirm our findings.

(J Med Internet Res 2021;23(5):e24346) doi:10.2196/24346

KEYWORDS

peripheral artery disease; fourth-generation synchronous telehealth program; ischemic stroke; hospitalization; cardiovascular disease; telehealth; stroke; cost; benefit; heart

Introduction

Cardiovascular (CV) diseases remain the biggest health burden worldwide [1]. Telemedicine can be used to monitor diseases and treat patients in real time. Previous studies have shown that patients with CV diseases who received telehealth medicine had better control over vascular risk factors, such as hypertension, diabetes mellitus, and dyslipidemia [2]. Furthermore, both Spyros et al [3] and Sally et al [4] reported that telemedicine was an important prognostic factor for patients with congestive heart failure in reducing all-cause mortality Relatedly, we have shown that a fourth-generation telehealth program—more specifically, an internet-based, synchronized disease management program that provides an immediate response—is associated with lower all-cause mortality among patients with CV as compared to that seen in patients who did not participate in the program [5]. However, Takahashi et al [6] found that asynchronous telemonitoring did not lead to fewer hospitalizations or emergency department visits. Meanwhile, a review article suggested that telemedicine should be carefully evaluated and applied in only those patients who would really benefit from it in terms of improved clinical outcomes [7].

Patients with peripheral artery disease (PAD) have relatively high rates of various comorbidities, including acute coronary syndrome (ACS), stroke, and congestive heart failure [8-10]. Numerous cohort studies have reported that the clinical outcomes of PAD patients are grave, with the rates of major adverse CV events and mortality being higher among PAD patients than among the general population. Moreover, PAD is an independent risk factor for mortality in patients with CV diseases [11]. Current guidelines advocate the use of antiplatelet agents and statins as primary therapies for these patients to prevent further major adverse cardiovascular events, because they are at high risk of CV events [11,12]. Therefore, it is important to treat PAD patients using innovative modalities to improve their overall prognosis. New treatment strategies that offer better CV outcomes for these patients are needed in the internet era. The costs and CV benefits of the fourth-generation synchronous telehealth program for these patients remain unknown.

The aims of this study were to elucidate the effects of the fourth-generation synchronous telehealth program for PAD patients on their CV outcomes, including stroke, ACS, and hospitalization for heart failure; and to evaluate the medical costs for these patients, including those who are treated at outpatient clinics and emergency departments or who are hospitalized.

Methods

Study Design

This was a retrospective cohort study in which data from a tertiary hospital in Taiwan was analyzed. This study was approved by the Institutional Review Board of National Taiwan University Hospital (NTUH), Taipei, Taiwan. All protocols were performed in accordance with relevant guidelines and regulations.

Patient Selection

Patients older than 20 years diagnosed with chronic CV diseases were screened at NTUH. Chronic CV diseases included coronary artery disease, myocardial infarction, heart failure, PAD, stroke, and hypertension. The patients with PAD were further selected for this study, with the PAD patients who participated in the fourth-generation synchronous telehealth program at the telehealth center of NTUH being enrolled as the study group. The patients with PAD who received usual care were enrolled as the control group. The decision of whether to participate in the telehealth program depended on the patients and/or their caregivers.

Telehealth Care Program

The fourth-generation synchronous telehealth program at our center is an integrated remote management program for chronic diseases. The internet-based platform was developed by the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taiwan. The details of this program have been reported previously [13]. Briefly, this telehealth program provides the following services: the collection of biometric data including single-lead electrocardiography, blood pressure, heart rate, and oximetry data, which are transferred from patients to our telehealth center daily and on demand; daily and on-demand telephone calls from nurse case managers for communication and health promotion; care and guidance from full-time nurse case managers and cardiologists who are in charge of care 24 hours per day; and long-term medication services and care management, which are discussed with the patients' primary care physician after acute events. The telehealth program bridges acute and home care and emphasizes education, prevention, and the early detection of clinical deterioration.

Usual Care

The patients in the control group received the usual care provided by the primary care physicians at our CV center according to updated guidelines including, but not limited to, the American Heart Association's guidelines for lifestyle

XSL•FO
RenderX

modification and primary prevention to reduce CV risks, guidelines for the management of stable ischemic heart disease, and the American Diabetes Association's guidelines for the management of diabetes. These patients made routine outpatient department visits (once every 3 months) to their primary care physicians. There was no contact between the telehealth center and the patients receiving usual care.

Covariates and Outcomes

The covariates were demographics (age, sex, and BMI), 10 comorbidities (hypertension, diabetes mellitus, hyperlipidemia, coronary artery disease, chronic kidney disease, myocardial infarction, stroke, heart failure, atrial fibrillation, and cancer), Charlson Comorbidity Index score, vital signs and laboratory data (mean arterial pressure, pulse rate, creatinine, and low-density lipoprotein-cholesterol), and the kinds of medication used at the time of enrollment. The outcomes were all-cause readmission, heart failure hospitalization, ACS, stroke, and the composite of ACS and stroke during follow-up. Each outcome was defined according to whether the patient was admitted with a principal discharge diagnosis during hospitalization.

Statistical Analysis

To reduce possible potential confounding when comparing the risk of outcomes between the telehealth and control groups, we employed an inverse probability of treatment weighting (IPTW) based on the propensity score [14]. The propensity score was estimated using a multivariable logistic regression model without consideration to interaction effects among the covariates, with the study group (1=telehealth, 0=control) being regressed on the selected covariates listed in Table 1 and the follow-up duration being replaced with the index date. We used the weight to estimate the average treatment effect and used a stabilized weight to mitigate the impact of extreme estimated propensity scores [15]. The balance of covariates between the groups before and after IPTW was checked using the absolute value of standardized difference between the groups, where a value less than 0.1 was considered a negligible difference [16].

The incidence of outcomes during follow-up was calculated using the incidence density ID expressed as number of events per 100 person-years. The absolute risk difference between the groups was also calculated in the IPTW-adjusted cohort. Furthermore, we compared the risk of outcomes during follow-up between the groups using the Cox proportional hazards model. The outcomes were assessed during the 1-year follow-up period, during the entire follow-up period, and before and after the 1-year follow-up period (landmark by 1 year). The average number of all-cause readmissions during follow-up between groups was compared between the groups using a Poisson model which treated the logarithm of follow-up duration as an offset variable. The accumulated medical expenditures during follow-up was compared between the groups using a linear regression model. The study group was the only explanatory factor in the aforementioned regression models. A 2-sided P value <.05 was considered statistically significant, and no adjustment of multiple testing (multiplicity) was made. All statistical analyses were performed using SAS version 9.4 (SAS Institute).

XSL•FO
RenderX

Table 1. Baseline characteristics of the study patients with peripheral arterial occlusive disease who participated (n=162) and did not participate (n=229) in the telemedicine program.

Variable	Before IPTW[a]			After IPTW		
	With[b] (n=162)	Without[c] (n=229)	STD[d]	With (n=162)	Without (n=229)	STD
Age (years), mean (SD)	70.8 (12.5)	70.0 (12.7)	0.06	70.1 (12.4)	70.2 (12.5)	–0.01
Male sex, n (%)	103 (63.6)	141 (61.6)	0.04	62.3	63.0	–0.01
Body mass index (kg/m^2), mean (SD[d]; n=374)	24.7 (4.4)	24.5 (4.0)	0.06	24.9 (4.6)	24.6 (3.9)	0.07
Comorbid conditions, n (%)						
Hypertension	107 (66.0)	148 (64.6)	0.03	66.1	64.9	0.03
Diabetes mellitus	68 (42.0)	100 (43.7)	–0.03	44.2	43.6	0.01
Hyperlipidemia	76 (46.9)	100 (43.7)	0.07	43.2	44.0	–0.02
Coronary artery disease	108 (66.7)	145 (63.3)	0.07	65.8	65.6	<0.01
Chronic kidney disease	49 (30.2)	60 (26.2)	0.09	27.1	28.9	–0.04
Old myocardial infarction	17 (10.5)	27 (11.8)	–0.04	10.6	11.9	–0.04
Stroke	53 (32.7)	65 (28.4)	0.09	29.9	29.8	<0.01
Heart failure	38 (23.5)	57 (24.9)	–0.03	22.4	24.1	–0.04
Atrial fibrillation	17 (10.5)	29 (12.7)	–0.07	14.5	11.6	0.09
Cancer	13 (8.0)	22 (9.6)	–0.06	8.7	9.0	–0.01
Charlson Comorbidity Index score, mean (SD)	3.1 (2.0)	3.0 (2.1)	0.05	3.0 (2.0)	3.1 (2.1)	–0.04
Laboratory data, mean (SD)						
Mean arterial pressure, (mmHg; n=354)	90.0 (14.2)	89.6 (14.3)	0.02	90.6 (13.6)	89.7 (14.1)	0.07
Pulse, (mmHg; n=354)	74.6 (12.9)	75.9 (14.0)	–0.10	75.8 (13.6)	75.8 (13.8)	<0.01
Creatinine, (mg/dL; n=385)	2.1 (2.8)	2.3 (2.6)	–0.07	2.1 (2.7)	2.4 (2.6)	–0.09
LDL–C[e] (mg/dL; n=318)	90.3 (28.2)	88.2 (28.8)	0.07	91.9 (27.4)	88.8 (29.1)	0.11
Medication, n (%)						
ACEI[f]/ARB[g]	82 (50.6)	121 (52.8)	–0.04	53.9	52.4	0.03
Beta blocker	96 (59.3)	129 (56.3)	0.06	60.7	58.5	0.05
dCCB[h]	87 (53.7)	115 (50.2)	0.07	51.3	52.4	–0.02
Diuretics	73 (45.1)	95 (41.5)	0.07	43.3	43.2	<0.01
Aspirin	81 (50.0)	125 (54.6)	–0.09	52.6	53.9	–0.02
Clopidogrel/Cilostazol	73 (45.1)	130 (56.8)	–0.24	51.3	51.9	–0.01
Oral anticoagulants	25 (15.4)	33 (14.4)	0.03	13.3	13.5	–0.01
Oral hypoglycemic agent	44 (27.2)	63 (27.5)	–0.01	27.1	27.9	–0.02
Insulin	12 (7.4)	29 (12.7)	–0.18	10.7	10.5	<0.01
Statin	70 (43.2)	97 (42.4)	0.02	39.3	41.9	–0.05
Follow-up period (years), mean (SD)	3.1 (1.8)	3.2 (1.8)	–0.09	3.3 (1.9)	3.1 (1.7)	0.14

[a]IPTW, inverse probability treatment weighting.

[b]Treated with the telemedicine program.

[c]Treated without the telemedicine program.

[d]STD: standardized difference.

[e]LDL-C: low-density lipoprotein-cholesterol.

[f]ACEi: angiotensin converting enzyme inhibitor.

[g]ARB: angiotensin receptor blocker.

XSL•FO
RenderX

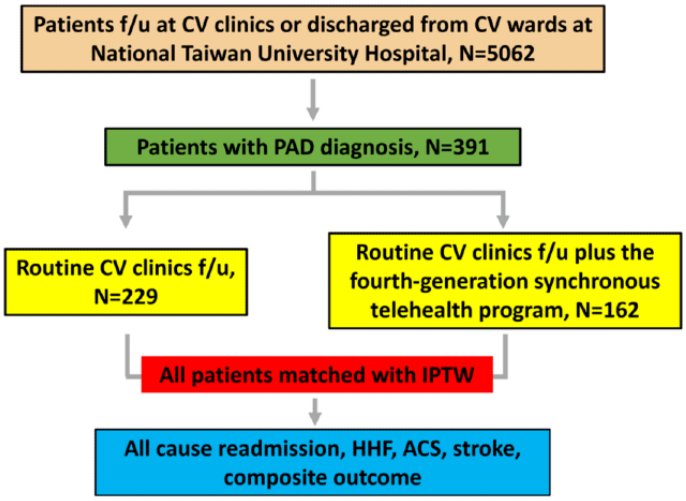
[h]dCCB: dihydropyridine calcium channel blocker.

Results

Patient Demographics and Clinical Features

A total of 5062 patients were screened in this study, and 391 patients with PAD diagnosis were further selected into the telehealth and control group (Figure 1). The baseline characteristics of those patients are reported in Table 1. In the telehealth groups, the mean age was 70.8 (SD 12.5) years, and 63.6% (103/162) of the patients were male. In the control group, the mean age was 70.0 (SD 12.7) years, and 61.6% (141/229) of the patients were male. The telehealth group had more patients with stroke (53/162, 32.7% vs 65/229, 28.4%, in the telehealth group vs control group, respectively), chronic kidney disease (49/162, 30.2% vs 60/229, 26.2%), coronary artery disease (108/162, 66.7% v. 145/229, 63.3%), and hyperlipidemia (76/162, 46.9% vs 100/229, 43.7%). The Charlson Comorbidity Index score was higher in the telehealth group (mean 3.1, SD 2.0) than in the control group (mean 3.0, SD 2.1). The mean follow-up time was 3.1 (SD 1.8) years for the telehealth group and 3.2 (SD 1.8) years for the control group. After propensity score weighting, the 2 study groups were found to have well-balanced characteristics (Table 1).

Figure 1. Flow chart of enrolled patients. ACS: acute coronary syndrome; CV: cardiovascular: f/u: follow-up; HHF: hospitalization for heart failure; IPTW: inverse probability of treatment weighting; PAD: peripheral artery disease.



Prognosis of Patients With PAD Participating in the Telehealth Program

During the 1-year follow-up period, 2 of the 162 (1.2%) patients in telehealth group and 8 of the 229 (3.5%) patients in the control group had stroke, respectively. The 1-year stroke rate was significantly lower in the telehealth group (0.9 vs 3.5 events per 100 person-years; hazard ratio [HR] 0.24; 95% CI 0.07-0.80) after IPTW. The incidence of composite vascular outcome in the telehealth cohort was also lower than that in the control cohort after IPTW (2.7 vs 5.8 events per 100 person-years; HR 0.46; 95% CI 0.21-0.997). However, no significant differences between the 2 were observed in terms of all-cause readmission, heart failure hospitalization, or ACS in the 1-year follow-up period. At the end of the follow-up period, only the stroke rate was significantly lower in the telehealth group (1.4 vs 2.6 events per 1000 person-years; HR 0.52; 95% CI 0.28-0.93) after IPTW (Table 2 and Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Follow-up outcomes (after inverse probability treatment weighting) in patients with peripheral arterial occlusive disease who participated (n=162) and did not participate (n=229) in the telemedicine program.

Outcome	Telemedicine event rate, %/ID[a] (95% CI)	Nontelemedicine, event rate, %/ID (95% CI)	ARD[b] (95% CI)	HR[c] of telemedicine (95% CI)	*P* value
One-year follow-up					
All-cause readmission	37.8/55.8 (46.8 to 64.8)	39.3/59.6 (50.2 to 69.0)	–0.38 (–1.68 to 0.92)	0.93 (0.75 to 1.17)	.55
HFH[d]	3.7/4.1 (2.3 to 6.8)	4.3/4.8 (2.8 to 7.7)	–0.71 (–3.84 to2.42)	0.85 (0.42 to 1.73)	.66
ACS[e]	1.6/1.7 (0.6 to 3.7)	2.3/2.5 (1.1 to 4.7)	–0.76 (–2.88 to 1.36)	0.69 (0.25 to 1.94)	.48
Stroke	0.9/0.9 (0.2 to 2.6)	3.5/3.9 (2.1 to 6.6)	–2.99 (–5.29 to 0.68)	0.24 (0.07 to 0.80)	.02
Composite outcome	2.4/ 2.7 (1.3 to 5.0)	5.2/ 5.8 (3.6 to 8.9)	–3.14 (–6.19 to 0.09)	0.46 (0.21 to 0.997)	.049
End of follow-up					
All-cause readmission	62.2/ 35.5 (31.2 to 40.3)	62.6/ 38.1 (33.4 to 43.1)	–2.5 (–9.1 to –4.0)	0.95 (0.79 to 1.13)	.56
HFH	10.1/ 3.1 (2.2 to 4.3)	7.6/ 2.6 (1.7 to 3.7)	0.55 (–0.79 to1.90)	1.25 (0.78 to 2.02)	.36
ACS	3.0/ 0.9 (0.5 to 1.6)	2.9/ 1.0 (0.5 to 1.7)	–0.04 (–0.80 to 0.73)	0.96 (0.43 to 2.17)	.92
Stroke	4.4/ 1.4 (0.8 to 2.2)	7.7/ 2.6 (1.8 to 3.7)	–1.24 (–2.37 to 0.11)	0.52 (0.28 to 0.93)	.03
Composite outcome	7.5/ 2.3 (1.6 to 3.4)	9.4/ 3.2 (2.3 to 4.4)	–0.86 (–2.20 to 0.48)	0.72 (0.44 to 1.17)	.18

[a]ID: incidence density.

[b]ARD: absolute risk difference.

[c]HR: hazard ratio.

[d]HHF: hospitalization for heart failure.

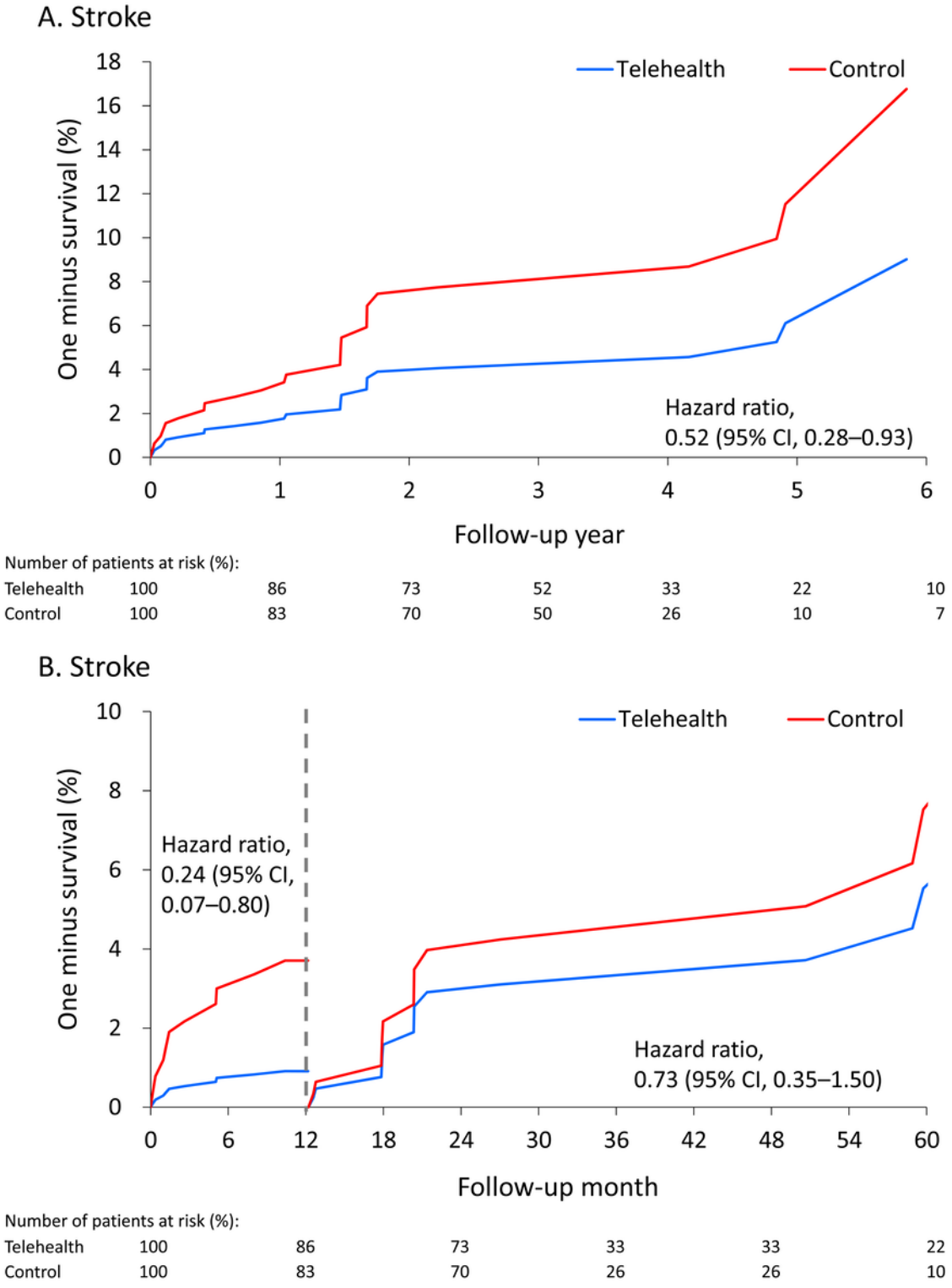
[e]ACS: acute coronary syndrome.

Landmark Analysis by 1-Year Follow-Up

The fitted (predicted) survival curve of stroke during the overall follow-up period was significantly lower in the telehealth group than in the control group (Figure 2A). The 2 curves remained significantly different during the 1-year follow-up period (HR 0.24; 95% CI 0.07-0.80) but were not significantly different afterwards (HR 0.73; 95% CI 0.35-1.50; Figure 2B). Similarly, the fitted (predicted) survival curve for composite vascular outcome (including acute coronary syndrome and stroke) during the overall follow-up period was not significantly lower in the telehealth group than in the control group (Figure 3A). The 2 curves were only significantly different in the 1-year follow-up after being stratified by 1-year follow-up (HR 0.46; 95% CI 0.21-0.997; Figure 3B).

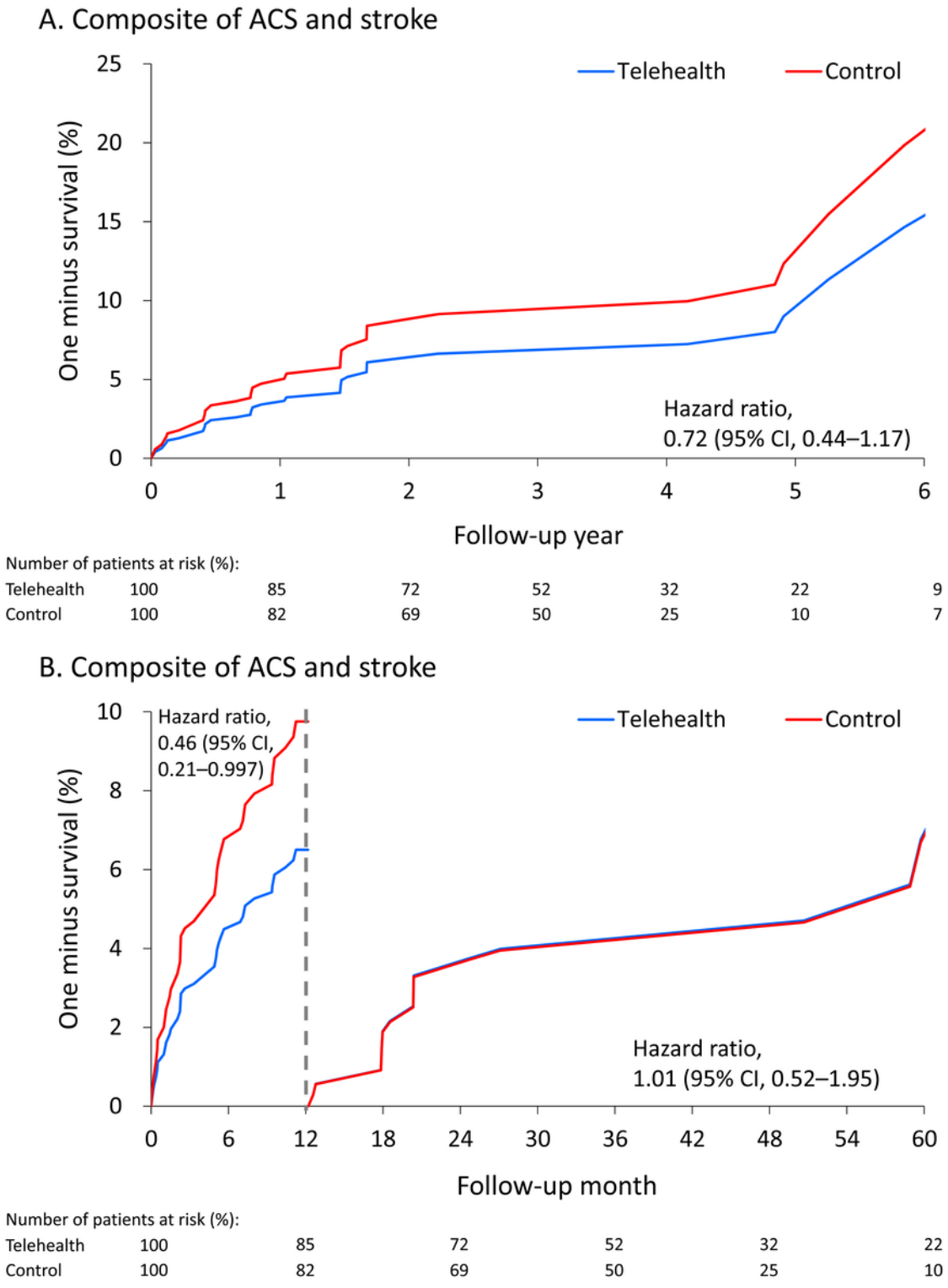
XSL•FO
RenderX

Figure 2. Fitted (predicted) survival curves of stroke during the overall follow-up (A) and stratified by 1-year follow-up (B) in patients who participated in the telemedicine program and who did not participate the telemedicine program.



XSL•FO
RenderX

Figure 3. Fitted (predicted) survival curves of composite arterial cardiovascular outcome during the overall follow-up (A) and stratified by 1-year follow-up (B) in patients who participated in the telemedicine program and who did not participate the telemedicine program.



Medical Utilization and Costs for Patients Participating in the Telehealth Program

The PAD patients in both the telehealth group and the control group had similar readmission times (mean 1.6, SD 2.0 vs mean 1.7, SD 2.2; rate ratio 1.08; 95% CI 0.97-1.20). Furthermore, no significant differences in medical expenditures between the 2 groups were observed in terms of outpatient, emergency department, hospitalization, or total costs during the whole course of follow-up (Table 3 and Multimedia Appendix 1).

XSL•FO
RenderX

Table 3. Medical utilization and cost (after inverse probability treatment weighting) in patients with peripheral arterial occlusive disease who participated (n=162) and did not participate (n=229) in the telemedicine program after inverse probability treatment weighting..

Outcome	Telemedicine, mean (SD)	Nontelemedicine, mean (SD)	RR[a] or *B*[b] of telemedicine (95% CI)	*P* value
Number of readmissions	1.8 (2.3)	1.6 (2.1)	1.08 (0.97 to 1.20)	.16
Medical expenditures (NTD[c]×10[3])				
Outpatient	577 (959)	528 (891)	577 (963)	.66
Emergency department	42 (55)	46 (63)	42 (54)	.32
Hospitalization	760 (853)	913 (1236)	819 (914)	.57
Total	1380 (1486)	1487 (1692)	1438 (1495)	.88

[a]RR: rate ratio.

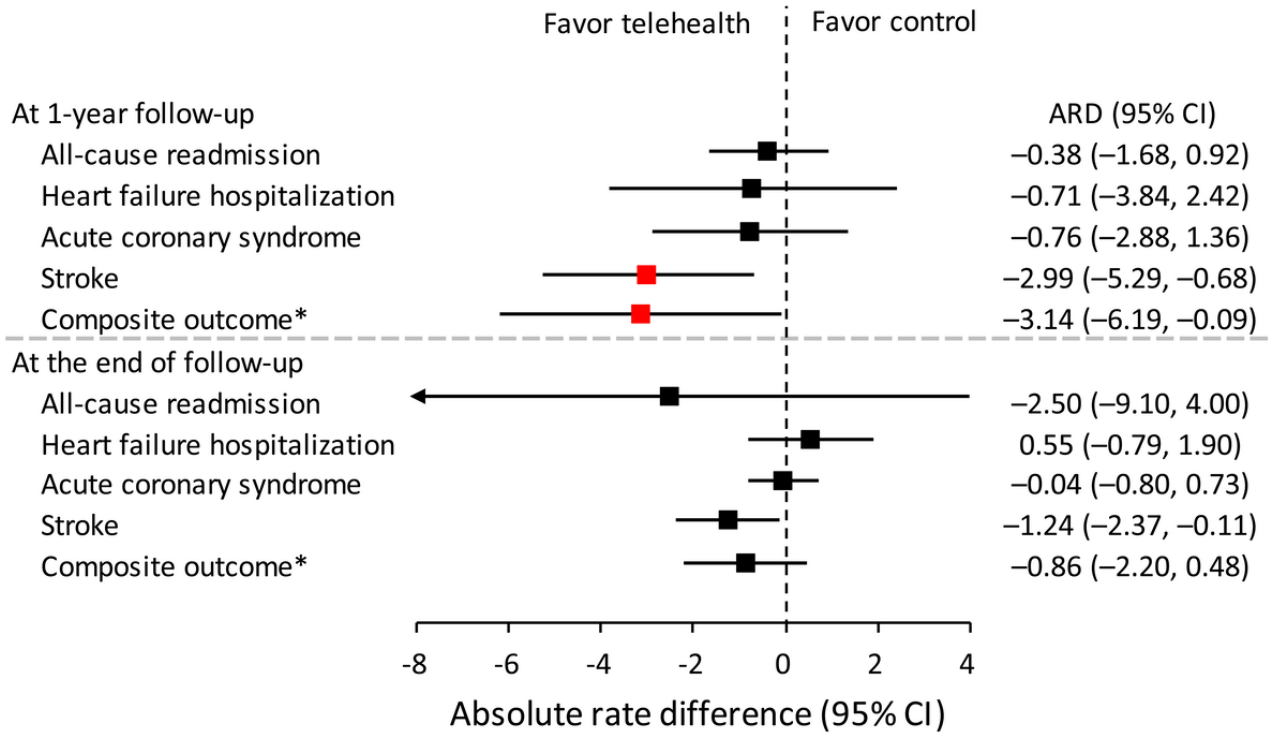
[b]*B*: regression coefficient.

[c]NTD: New Taiwan dollar.

Incidence Rate of All Outcomes in Patients Who Participated in the Telehealth Program

The absolute risk difference in the incidence (expressed as number of events per 100 person-years) of stroke and arterial vascular composite outcome was –2.99 (95% CI –5.29 to –0.68) and –3.14 (95% CI –6.19 to –0.09), respectively, at the 1-year follow-up. At the end of the follow-up period, only the absolute risk difference of the incidence rate for stroke was significant (Figure 4).

Figure 4. The difference of the incidence rate (expressed as number of event per 100 person-years) of all outcomes in patients who participated in the telemedicine program and who did not participate in the inverse probability of treatment weighting–adjusted cohort. ACS: acute coronary syndrome; ARD: absolute risk difference. *Composite of ACS and stroke.



Discussion

This is the first study to investigate the application of the fourth-generation synchronous telehealth program to patients with PAD in terms of various CV outcomes. Retrospectively, we found that the PAD patients who participated in the telehealth program had lower risks of both stroke and composite vascular events in the midterm follow-up but only a lower risk of stroke events over the long-term follow-up period. The vascular protection seemed most prominent in the first year, and this phenomenon then diminished in the following years with regular CV clinic visits. In the first year, 3 out of every 100 patients in the telemedicine group were free from stroke events. Otherwise, the total medical costs were comparable in both groups.

Telehealth care has been shown to reduce hospitalizations in patients with chronic conditions such as asthma, chronic obstructive pulmonary disease, and heart failure [17-20]. We have previously reported that better cost-effectiveness and clinical outcomes were noted with the use of fourth-generation

XSL•FO
RenderX

synchronous telehealth program in patients with chronic CV diseases [13]. However, some other studies failed to show better clinical outcomes in patients taking part in telehealth programs [6,7]. This difference in the results could be explained by the enrollment of different types of patients and the implementation of different versions of telehealth programs. We previously reported that PAD was the remaining prognostic factor for CV admission in patients receiving telehealth care after multivariable Cox regression [21]. However, the costs and CV benefits of the fourth-generation synchronous telehealth program for PAD patients are still not well validated. In this study, we focused on those patients with PAD who had higher risks of grave CV outcomes such as myocardial infarction and stroke [22,23]. According to a previous study, the 5-year cumulative incidence of mortality among asymptomatic and symptomatic patients with PAD is 19% and 27%, respectively [24]. In the internet era, it is important to consider the speed of medical care responses of telehealth programs, especially with respect to high-risk patient groups. The early identification of PAD patients with optimized and comprehensive treatment is critical for improving clinical outcomes, and telehealth programs could be applied to PAD-related wound care or gangrene management in place of CV systems that provide comprehensive care [25,26]. In this study, we found that patients receiving telemedicine care had significantly lower rates of stroke events in the mid- and long-term follow-up periods than did the group that received normal care. As high blood pressure is a major cause of stroke, better control of blood pressure is essential to preventing stroke attacks. Lu et al [27] showed that the implementation of home telehealth care supervised by nursing managers improved blood pressure control in patients with high blood pressure. Furthermore, it was found that a mobile phone intervention could significantly reduce systolic blood pressure among patients with high blood pressure (mean difference of 4.3 mmHg) [28]. Digital health strategies were also found to have a beneficial effect on blood pressure control [29]. These findings suggest that the stroke prevention benefits of telemedicine may be linked to the better blood pressure control exhibited by patients using the telehealth program. On the other hand, the prevalence of carotid artery stenosis is high in patients with PAD [30]. Through good medication compliance, the early management of CV risk factors, and the prompt detection of clinical arrhythmia, full-time nursing case managers and cardiologists who provide round-the-clock care are also able to reduce the risk of stroke among those patients with PAD that have carotid artery stenosis.

In this study, we found that, in the midterm, the fourth-generation synchronous telehealth programs improved the composite vascular outcome (including acute coronary syndrome and stroke) of patients with PAD. This vascular protection was most prominent in the first year and gradually diminished over the years that followed as confirmed through the regular follow-up visits at CV clinics. The telehealth program seems to bridge the transition from hospital admission to home care. Therefore, the proper timing for initiating telehealth is just as important as the timing for taking medication. The conditions of patients with CV disease deteriorated rapidly just after their discharge [31,32]. Our study supports the notion that the optimal timing for implementing a telehealth program for patients with PAD is at the time of discharge from hospital.

Those patients with PAD who took part in the telehealth program in this study had 2.99 fewer stroke events and 1.86 fewer composite vascular events per 100 patients in the first year than did the control group patients. As the prevalence and incidence of patients with PAD is high, the number of patients needed to be treated with telehealth program can be easily achieved in clinical practice. As patients participating in telehealth programs will have to pay additional fees, the costs and benefits associated with this program should be analyzed. We found that the medical costs of the telehealth program patients in this study were not greater than those of the control group, in spite of the fact that they did have better clinical outcomes in terms of stroke and composite vascular events. This finding suggests that the participation of PAD patients in the telehealth programs can also confer economic benefits. Moreover, opportunity costs other than medical ones should also be considered. Because patients with PAD suffer from grave CV outcomes, they are more likely to lose their jobs and to pay higher insurance premiums. In the United States, individuals diagnosed with PAD face higher health care–related expenditures and out-of-pocket expenses compared to healthy adults [33]. Therefore, the total expenditure of patients with PAD and stroke is likely higher than that revealed in our study.

This study had several limitations. First, the study was not randomized, which resulted in the heterogeneity of the patient population and disease severity. However, we performed IPTW matching of 2 groups with balance to minimize the possible confounding effect of clinical factors. Second, the outcome of amputation was not coded in our cohort. In addition, although we did show the composite vascular outcome in our analysis, this outcome is particularly important in patients with PAD and should be clarified in a future study. Third, the clinical outcomes were derived from the electronic billing and medical records of our hospital, and the data for patients who received care outside of our hospital were not recorded. Resources that were used but not billed might have also been overlooked when extracting data from our billing system.

Those patients with PAD who participated in the fourth-generation telehealth program exhibited a lower risk of stroke events over both short- and long-term follow-up periods than did the control group. In addition, the total medical costs of the 2 groups were comparable, with no significant difference being found. Further randomized trials are needed, however, to confirm our findings and guide clinical practice.

Acknowledgments

We thank the staff of the Department of Medical Research, NTUH for the Integrated Medical Database. We also thank Mr Alfred Hsing-Fen Lin for his assistance with the statistical analysis during the completion of the manuscript.

XSL•FO
RenderX

This research was supported by the Ministry of Science and Technology of Taiwan (no. MOST 108-2221-E-002-163- and MOST 109-2221-E-002 -083 -) and NTUH (no. 107-EDN11, 108-N4406, 108EDN02, 109-O20, 109-S4579, and 109-EDN11).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Follow-up outcomes in patients with peripheral arterial occlusive disease who participated and did not participate in the telemedicine program; and medical utilization and cost in patients with peripheral arterial occlusive disease who participated and did not participate in the telemedicine program.
[DOCX File , 32 KB - jmir_v23i5e24346_app1.docx]

References

1. Kung H, Hoyert D, Xu J, Murphy S. Deaths: final data for 2005. Natl Vital Stat Rep Apr 2008;56(10):1-120. [doi: 10.1037/e518032006-001]
2. Vernooij JWP, Kaasjager HAH, van der Graaf Y, Wierdsma J, Grandjean HMH, Hovens MMC, SMARTStudy Group. Internet based vascular risk factor management for patients with clinically manifest vascular disease: randomised controlled trial. BMJ 2012 Jun 12;344:e3750 [FREE Full text] [doi: 10.1136/bmj.e3750] [Medline: 22692651]
3. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
4. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040 [FREE Full text] [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
5. Hung C, Yu J, Lin Y, Chen Y, Huang C, Lee J, et al. Mortality benefit of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease: a longitudinal study. J Med Internet Res 2016 May 13;18(5):e102 [FREE Full text] [doi: 10.2196/jmir.5718] [Medline: 27177497]
6. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
7. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
8. Hedberg P, Hammar C, Selmeryd J, Viklund J, Leppert J, Hellberg A, et al. Left ventricular systolic dysfunction in outpatients with peripheral atherosclerotic vascular disease: prevalence and association with location of arterial disease. Eur J Heart Fail 2014 Jun;16(6):625-632 [FREE Full text] [doi: 10.1002/ejhf.95] [Medline: 24771615]
9. Banerjee A, Fowkes FG, Rothwell PM. Associations between peripheral artery disease and ischemic stroke. Stroke 2010 Sep;41(9):2102-2107. [doi: 10.1161/strokeaha.110.582627]
10. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. Circ Res 2015 Apr 24;116(9):1509-1526. [doi: 10.1161/circresaha.116.303849]
11. Gerhard-Herman M, Gornik H, Barrett C, Barshes N, Corriere M, Drachman D, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Circulation 2017 Mar 21;135(12):E. [doi: 10.1161/cir.0000000000000470]
12. Aboyans V, Ricco J, Bartelink M, Björck M, Brodmann M, Cohnert T, et al. 2017 ESC guidelines on the diagnosis and treatment of peripheral arterial diseases, in collaboration with the European Society for Vascular Surgery (ESVS). Eur J Vasc Endovasc Surg Mar 2018;55(3):305-368. [doi: 10.5603/kp.2017.0216]
13. Ho Y, Yu J, Lin Y, Chen Y, Huang C, Hsu T, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. J Med Internet Res 2014 Jun 10;16(6):e145 [FREE Full text] [doi: 10.2196/jmir.3346] [Medline: 24915187]
14. 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]
15. 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]

XSL•FO
RenderX

16. Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly. Journal of Clinical Epidemiology 2001 Apr;54(4):387-398. [doi: 10.1016/s0895-4356(00)00321-8]
17. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J. Telehealthcare for asthma. Cochrane Database Syst Rev 2010 Oct;10:17. [doi: 10.1002/14651858.cd007717.pub2]
18. McLean S, Nurmatov U, Liu J, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011;06(7):18. [doi: 10.1002/14651858.cd007718.pub2]
19. Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 05;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
20. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
21. Lee J, Hung C, Huang C, Chen Y, Chuang P, Yu J, et al. Use of the CHA2DS2-VASc score for risk stratification of hospital admissions among patients with cardiovascular diseases receiving a fourth-generation synchronous telehealth program: retrospective cohort study. J Med Internet Res 2019 Jan 31;21(1):e12790 [FREE Full text] [doi: 10.2196/12790] [Medline: 30702437]
22. Olin JW, Sealove BA. Peripheral artery disease: current insight into the disease and its diagnosis and management. Mayo Clin Proc 2010 Jul;85(7):678-692 [FREE Full text] [doi: 10.4065/mcp.2010.0133] [Medline: 20592174]
23. Shigematsu H. Three-year cardiovascular events and disease progress in patients with peripheral arterial disease: results from the Japan Medication Therapy for Peripheral Arterial Disease (J-METHOD). Int Angiol 2010 Apr;29(2 Suppl) 2010 Apr 29:2-13. [doi: 10.3109/9781420019667-23]
24. Sigvant B, Lundin F, Wahlberg E. The risk of disease progression in peripheral arterial disease is higher than expected: a meta-analysis of mortality and disease progression in peripheral arterial disease. Eur J Vasc Endovasc Surg 2016 Mar;51(3):395-403 [FREE Full text] [doi: 10.1016/j.ejvs.2015.10.022] [Medline: 26777541]
25. Boodoo C, Perry JA, Hunter PJ, Duta DI, Newhook SCP, Leung G, et al. Views of patients on using mHealth to monitor and prevent diabetic foot ulcers: qualitative study. JMIR Diabetes 2017 Sep 15;2(2):e22 [FREE Full text] [doi: 10.2196/diabetes.8505] [Medline: 30291089]
26. Wallace D, Perry J, Yu J, Mehta J, Hunter P, Cross KM. Assessing the need for mobile health (mHealth) in monitoring the diabetic lower extremity. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11879 [FREE Full text] [doi: 10.2196/11879] [Medline: 30990455]
27. Lu J, Chen C, Hsu C. Effect of home telehealth care on blood pressure control: A public healthcare centre model. J Telemed Telecare 2017 Oct 17;25(1):35-45. [doi: 10.1177/1357633x17734258]
28. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e16695 [FREE Full text] [doi: 10.2196/16695] [Medline: 32628615]
29. Pellegrini D, Torlasco C, Ochoa JE, Parati G. Contribution of telemedicine and information technology to hypertension control. Hypertens Res 2020 Jul;43(7):621-628. [doi: 10.1038/s41440-020-0422-4] [Medline: 32203451]
30. Hyuk JJ, Sang SL, Hyun YK, Byung SP, Dong IK, Kyoung JN, et al. Association between carotid artery stenosis and peripheral artery diseasevaluation by screening carotid ultrasonography (cross-sectional study). Medicine (Baltimore) Feb 2019;98(6):E. [doi: 10.1097/md.0000000000014163]
31. Desai AS. The three-phase terrain of heart failure readmissions. Circ Heart Fail 2012 Jul;5(4):398-400. [doi: 10.1161/circheartfailure.112.968735]
32. Chun S, Tu JV, Wijeysundera HC, Austin PC, Wang X, Levy D, et al. Lifetime analysis of hospitalizations and survival of patients newly admitted with heart failure. Circ Heart Fail 2012 Jul 01;5(4):414-421 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.111.964791] [Medline: 22556322]
33. Scully RE, Arnaoutakis DJ, DeBord Smith A, Semel M, Nguyen LL. Estimated annual health care expenditures in individuals with peripheral arterial disease. J Vasc Surg 2018 Feb;67(2):558-567 [FREE Full text] [doi: 10.1016/j.jvs.2017.06.102] [Medline: 28847660]

Abbreviations

ACS: acute coronary syndrome
CV: cardiovascular
HR: hazard ratio
IPTW: inverse probability of treatment weighting
NTUH: National Taiwan University Hospital
PAD: peripheral artery disease

XSL•FO
RenderX

Edited by G Eysenbach; submitted 24.09.20; peer-reviewed by M Singh, N Gronewold; comments to author 16.10.20; revised version received 22.10.20; accepted 14.04.21; published 18.05.21.

Please cite as:
Lee JK, Hung CS, Huang CC, Chen YH, Wu HW, Chuang PY, Yu JY, Ho YL
The Costs and Cardiovascular Benefits in Patients With Peripheral Artery Disease From a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study
J Med Internet Res 2021;23(5):e24346
URL: https://www.jmir.org/2021/5/e24346
doi:10.2196/24346
PMID:34003132

©Jen-Kuang Lee, Chi-Sheng Hung, Ching-Chang Huang, Ying-Hsien Chen, Hui-Wen Wu, Pao-Yu Chuang, Jiun-Yu Yu, Yi-Lwun Ho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.05.2021. 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 Impact of Inpatient Telemedicine on Personal Protective Equipment Savings During the COVID-19 Pandemic: Cross-sectional Study

Reem Halabi[1], BSc; Geoffrey Smith[2], MPS; Marc Sylwestrzak[2], BS; Brian Clay[2,3,4], MD; Christopher A Longhurst[2,3,5], MS, MD; Lina Lander[2,6], ScD

[1]School of Medicine, University of California, San Diego, San Diego, CA, United States
[2]Information Services, University of California San Diego Health, San Diego, CA, United States
[3]Health Sciences Department of Biomedical Informatics, University of California, San Diego, San Diego, CA, United States
[4]Department of Medicine, University of California, San Diego, San Diego, CA, United States
[5]Department of Pediatrics, University of California, San Diego, San Diego, CA, United States
[6]Department of Family Medicine, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Lina Lander, ScD
Department of Family Medicine
University of California, San Diego
9500 Gilman Drive MC0606
La Jolla, CA, 92093-0606
United States
Phone: 1 8582601917
Email: linalander@health.ucsd.edu

Abstract

With the emergence of the COVID-19 pandemic and shortage of adequate personal protective equipment (PPE), hospitals implemented inpatient telemedicine measures to ensure operational readiness and a safe working environment for clinicians. The utility and sustainability of inpatient telemedicine initiatives need to be evaluated as the number of COVID-19 inpatients is expected to continue declining. In this viewpoint, we describe the use of a rapidly deployed inpatient telemedicine workflow at a large academic medical center and discuss the potential impact on PPE savings. In early 2020, videoconferencing software was installed on patient bedside iPads at two academic medical center teaching hospitals. An internal website allowed providers to initiate video calls with patients in any patient room with an activated iPad, including both COVID-19 and non–COVID-19 patients. Patients were encouraged to use telemedicine technology to connect with loved ones via native apps or videoconferencing software. We evaluated the use of telemedicine technology on patients' bedside iPads by monitoring traffic to the internal website. Between May 2020 and March 2021, there were a total of 1240 active users of the Video Visits website (mean 112.7, SD 49.0 connection events per month). Of these, 133 (10.7%) connections were made. Patients initiated 63 (47.4%) video calls with family or friends and sent 37 (27.8%) emails with videoconference connection instructions. Providers initiated a total of 33 (24.8%) video calls with the majority of calls initiated in August (n=22, 67%). There was a low level of adoption of inpatient telemedicine capability by providers and patients. With sufficient availability of PPE, inpatient providers did not find a frequent need to use the bedside telemedicine technology, despite a high census of patients with COVID-19. Compared to providers, patients used videoconferencing capabilities more frequently in September and October 2020. We did not find savings of PPE associated with the use of inpatient telemedicine.

(J Med Internet Res 2021;23(5):e28845) doi:10.2196/28845

KEYWORDS

inpatient telemedicine; bedside iPad; video visits; personal protective equipment; COVID-19; virtual visits; pandemic; telehealth; telemedicine; digital health

XSL•FO
RenderX

Expansion of Telemedicine and Patient iPad Program

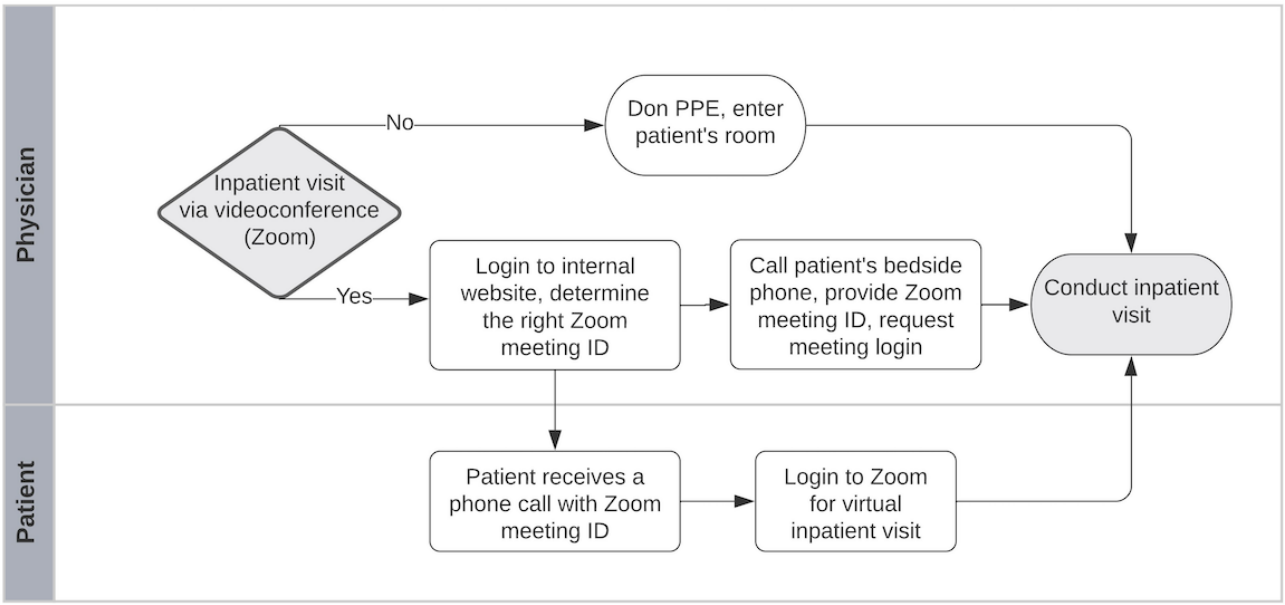
In response to the COVID-19 pandemic, health care organizations rapidly implemented telemedicine protocols to protect health care workers and patients [1-3]. The anticipated surge in patients and shortage in personal protective equipment (PPE) were key drivers for the implementation of virtual visits and medical screening exams in the inpatient settings and emergency departments [4]. Some authors referred to the inpatient telemedicine as "electronic PPE" [4], and a few studies saw a decrease in patient contact, resulting in PPE savings [5,6]. One study among emergency department patients found telemedicine to be associated with a significant reduction in PPE use, without anticipated increase in anxiety and dissatisfaction among patients [6]. Overcoming operational barriers such as privacy, device availability, and functionality may enable organizations to conserve PPE while providing excellent patient care [7]. Although a few studies described the deployment of hospital tablets for inpatient telemedicine, the extent of long-term continuous use of this technology has not been extensively evaluated [8,9].

University of California (UC) San Diego Health is an academic medical center with multiple outpatient ambulatory sites and two hospital campuses with 808 licensed inpatient beds. Bedside iPads were implemented at one affiliated teaching hospital in 2016 followed by expansion in 2019 to include the entire inpatient population [10]. The original intent of the bedside iPads was to allow patients access to their inpatient electronic health record portal (Epic MyChart Bedside), hospitality (ordering food), comfort (controlling ambient lights and room temperature), and entertainment (streaming services) [10]. As the COVID-19 pandemic evolved and more patients required isolation precautions, hospitals limited visitations by family members. Bedside iPads, therefore, appeared to be a natural solution to connect patients with their loved ones and to conserve PPE by allowing clinicians to interact with patients remotely. The objective of this viewpoint is to describe the use of a rapidly deployed inpatient telemedicine workflow at a large academic medical center and discuss technology's potential impact on PPE use.

Inpatient Telemedicine Process Implementation

A videoconferencing tool (Zoom, Zoom Video Communications) was implemented with health care security protocols in 2019 prior to the COVID-19 pandemic. The native Zoom app was installed on all the bedside iPads to allow patients to connect with their families using video calls. Inpatient providers could also initiate Zoom calls with patients. At the beginning of the pandemic, providers could call the patient's bedside telephone to provide them with a meeting ID to join a Zoom videoconference. The patient would then join the meeting initiated by the provider (Figure 1).

Figure 1. Process flow for inpatient video call at an academic medical center in San Diego, CA. PPE: personal protective equipment.



Although this process was adopted by many inpatient services, providers reported difficulties with retrieving telephone numbers for inpatient rooms. As a result, an internal website was created to display all available active patient iPads. From this website, providers could initiate a video call with a patient using a "Meet Now" button. The bedside iPad would then produce an audible ping alerting the patient of an incoming Zoom call. Upon discharge, the bedside iPad and Zoom ID were reset to ensure patient data privacy.

To evaluate the use of video calls in inpatient services and estimate the potential impact on PPE savings, we obtained data from the internal web server logs and Google analytics. Data from video calls to patients hospitalized for COVID-19 and other conditions were included. All available traffic data from the internal website were collected using internal servers between May and June 2020, before switching to using Google Analytics in July 2020.

XSL•FO
RenderX

We used Google Analytics to track user engagement events on patients' iPads. Engagement events included clicking on the Start Meeting, Send Email, and Meet Now links. The Start Meeting button allowed a patient to initiate a video call with their family members or a provider and was used as an indicator for a patient-initiated video call. The Send Email button allowed a patient to email video call instructions to their family members and was used as another indicator of a patient-initiated call. The Meet Now button allowed a physician to initiate a video call with a patient and was used as an indicator for provider-initiated video calls.

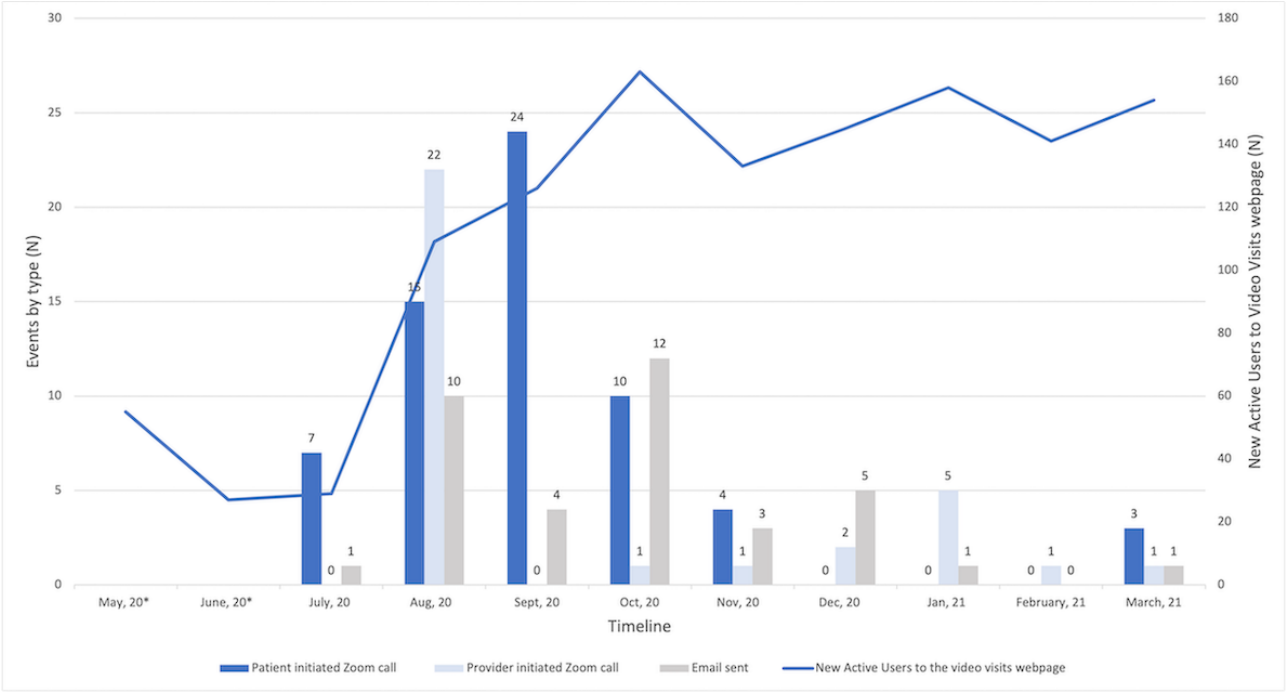
It was not technically feasible to track the volume of telephone calls from providers to patient rooms to determine the frequency of phone or video visits during the initial stage of Zoom implementation. It was also not feasible to track Zoom meetings conducted through the native Zoom app due to UC San Diego Health cybersecurity protocols.

We did not collect any patient identifying information and followed all data safety protocols. We relied on server and Google analytics data to determine whether patients used the bedside iPads to participate in a video call or to send an email, not the duration or content of such engagements.

Providers and Patients Underused Inpatient Telemedicine

The number of active users on the Video Visits website was determined from the internal server data for May and June 2020, and from Google Analytics for July 2020 to March 2021. There were a total of 1240 events collected from both data sources. On average, between May 2020 and March 2021, there were 112.73 (SD 49) connection events per month (Figure 2). As expected, use increased between July (n=29) and March (n=154). The lowest use was recorded in June (n=27) and the highest in October (n=163), followed by January 2021 with 158 users (Figure 2, line).

Figure 2. Users of the Video Visits webpage and engagement events (patient- and provider-initiated calls, emails), May 2020-March 2021, San Diego, CA. *The number of users of the Video Visits website was determined from the internal server data for May-June 2020 and from Google Analytics for July 2020-March 2021.



Although the number of visitors to the Video Visits site was low, the actual engagement with the system was even lower with 133 (10.7%) connections out of the total 1240 events. Of these 133 connections, patients initiated 63 (47.4%) video calls with family or friends and sent 37 (27.8%) emails with connection instructions. The highest number of calls occurred in September 2020 (n=24, 38%), and the largest number of emails were sent in October (n=12, 32%). Provider use was even lower, with providers initiating a total of 33 (52%) video calls with patients. The majority of calls were initiated in August (n=22, 67%) perhaps due to an internal awareness campaign of the available capability. Because the servers filter out automatic page refreshes, these numbers may underestimate the total number of video calls.

Inpatient Telemedicine Could Be Used to Conserve PPE

Although we implemented convenient workflows for clinicians to contact patients virtually to reduce the risk for infection, save PPE, and potentially save time to don and doff PPE, we saw minimal adoption of inpatient telemedicine. There appeared to be some seasonal variation in use, and during our study, patients initiated higher number of calls than providers (63 vs 33). The majority of our provider calls occurred in 1 month (August 2020) possibly due to an internal awareness campaign in response to a COVID-19 wave occurring at the time. Use declined to 0 to 5 calls in subsequent months, however. Our results are different from those of previous studies that found

XSL•FO
RenderX

high inpatient telemedicine use and subsequent savings of PPE [4-6,8].

No technical issues with iPads or Zoom were reported, yet the vast majority of providers saw patients in person. Inpatient physicians may prefer to see patients in person, regardless of availability and convenience of technology or potential saving of PPE. Fortunately, our academic medical center has not experienced the PPE shortages faced by our colleagues in other parts of the country [11]. As such, there was never a significant impetus to conserve PPE using all available means including inpatient telemedicine capability.

Our study had several limitations including small sample size and limited scope. Although most clinicians were comfortable with Zoom, it is possible that some were not, which would reduce technology use for inpatient visits. Patients have various comfort levels with technology, which could reduce the potential usability of inpatient telemedicine. In addition, we did not evaluate the utility of inpatient telemedicine for other health care providers such as nurses and technicians.

With the COVID-19 pandemic, telemedicine capacity was rapidly implemented in ambulatory settings [1-3]. Emergency departments also found multiple use cases for telemedicine including patient triage and expedited care for stable patients, reduced potential provider exposure to COVID-19, decreased use of PPE, reducing patient isolation, and allowing quarantined physicians to continue practicing, among others [4-6,12,13].

The utility of telemedicine in inpatient settings and its potential impact on PPE savings, however, remains to be addressed. A few studies that found inpatient telemedicine to be useful for patient care and PPE savings also reported operational barriers such as privacy and security, and limited functionality and device availability [7-9]. The availability of the technology by itself was inadequate to encourage use while PPE supplies were sufficient, and additional studies are needed to evaluate the utility of inpatient telemedicine for health care providers.

Conclusions

Similar to outpatient clinics and emergency departments, telemedicine capability could have a positive impact in inpatient settings including PPE savings. Our study found low adoption levels of inpatient telemedicine among patients and providers. Despite a high census of patients with COVID-19, our providers saw patients in person rather than relying on telemedicine. Patients' use of bedside iPads appeared to be limited for telemedicine and should continue to be evaluated to improve patient experience.

Increasing awareness of telemedicine in inpatient settings for providers and patients could help increase use for patient care. Going forward, implementation of telemedicine across all levels of care may help reduce potential provider exposure to COVID-19 and increase provider capacity.

Acknowledgments

The authors thank the UC San Diego Health Information Services Department teams for rapid implementation of inpatient telemedicine capability. Thank you to our patients and physicians who used inpatient telemedicine capability.

Authors' Contributions

LL and CAL designed the study. RH, GS, and MS obtained primary data. All authors contributed to data interpretation. RH and LL wrote the manuscript. All authors critically reviewed the manuscript for important intellectual content and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Hron JD, Parsons CR, Williams LA, Harper MB, Bourgeois FC. Rapid implementation of an inpatient telehealth program during the COVID-19 pandemic. Appl Clin Inform 2020 May;11(3):452-459. [doi: 10.1055/s-0040-1713635] [Medline: 32610350]
2. Meyer BC, Friedman LS, Payne K, Moore L, Cressler J, Holberg S, et al. Medical undistancing through telemedicine: a model enabling rapid telemedicine deployment in an academic health center during the COVID-19 pandemic. Telemed J E Health 2020 Oct 07:1. [doi: 10.1089/tmj.2020.0327] [Medline: 33030985]
3. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
4. Turer RW, Jones I, Rosenbloom ST, 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]
5. Lowe J, Shen S. Off the shelf: rapid deployment of an emergency department telemedicine platform using readily available consumer products. J Emerg Med 2020 Nov;59(5):726-729 [FREE Full text] [doi: 10.1016/j.jemermed.2020.09.026] [Medline: 33097352]

XSL•FO
RenderX

6. Candel BGJ, Vaes SMM, van Bree EM, Groenen SMA, Derkx F, Mignot LAA, et al. Telemedicine in the emergency department to decrease personal protective equipment use: a before-and-after study. Emerg Med J 2021 Mar;38(3):224-228. [doi: 10.1136/emermed-2020-210292] [Medline: 33355305]
7. Ong SY, Stump L, Zawalich M, Edwards L, Stanton G, Matthews M, et al. Inpatient telehealth tools to enhance communication and decrease personal protective equipment consumption during disaster situations: a case study during the COVID-19 pandemic. Appl Clin Inform 2020 Oct;11(5):733-741. [doi: 10.1055/s-0040-1719180] [Medline: 33147644]
8. Jones MS, Goley AL, Alexander BE, Keller SB, Caldwell MM, Buse JB. Inpatient transition to virtual care during COVID-19 pandemic. Diabetes Technol Ther 2020 Jun;22(6):444-448. [doi: 10.1089/dia.2020.0206] [Medline: 32396395]
9. 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]
10. Tai-Seale M, Downing NL, Jones VG, Milani RV, Zhao B, Clay B, et al. Technology-enabled consumer engagement: promising practices at four health care delivery organizations. Health Aff (Millwood) 2019 Mar;38(3):383-390. [doi: 10.1377/hlthaff.2018.05027] [Medline: 30830826]
11. He S, Ojo A, Beckman AL, 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]
12. Tresenriter M, Holdaway J, Killeen J, Chan T, Dameff C. The implementation of an emergency medicine telehealth system during a pandemic. J Emerg Med 2021 Apr;60(4):548-553 [FREE Full text] [doi: 10.1016/j.jemermed.2020.11.026] [Medline: 33423835]
13. Bains J, Greenwald PW, Mulcare MR, Leyden D, Kim J, Shemesh AJ, et al. Utilizing telemedicine in a novel approach to COVID-19 management and patient experience in the emergency department. Telemed J E Health 2021 Mar;27(3):254-260. [doi: 10.1089/tmj.2020.0162] [Medline: 32821027]

Abbreviations

PPE: personal protective equipment
UC: University of California

Edited by C Basch; submitted 16.03.21; peer-reviewed by E van der Velde, M Elhadi; comments to author 04.04.21; revised version received 20.04.21; accepted 01.05.21; published 19.05.21.

Please cite as:
Halabi R, Smith G, Sylwestrzak M, Clay B, Longhurst CA, Lander L
The Impact of Inpatient Telemedicine on Personal Protective Equipment Savings During the COVID-19 Pandemic: Cross-sectional Study
J Med Internet Res 2021;23(5):e28845
URL: https://www.jmir.org/2021/5/e28845
doi:10.2196/28845
PMID:33945494

©Reem Halabi, Geoffrey Smith, Marc Sylwestrzak, Brian Clay, Christopher A Longhurst, Lina Lander. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Characteristics of Citizens and Their Use of Teleconsultations in Primary Care in the Catalan Public Health System Before and During the COVID-19 Pandemic: Retrospective Descriptive Cross-sectional Study

Oscar Solans[1,2], MSc, MD; Josep Vidal-Alaball[3,4,5], MD, MPH, PhD; Pasqual Roig Cabo[6], MD; Núria Mora[7], BSc; Ermengol Coma[7], MD, MPH; Josep Maria Bonet Simó[6], MD; Eduardo Hermosilla Pérez[7,8], BSc; Francesc Saigí-Rubió[9,10], MD, PhD; Carmen Olmos Domínguez[1], MSc, MD; Jordi Piera-Jiménez[2,11,12], MSc; Mercè Abizanda González[13,14], BSc; Francesc López Seguí[6,15], PhD

[1]Health Department, Catalan Ministry of Health, Barcelona, Spain

[2]Digitalization for the Sustainability of the Healthcare System, Sistema de Salut de Catalunya, Barcelona, Spain

[3]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[4]Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

[5]Faculty of Medicine, University of Vic - Central University of Catalonia, Vic, Spain

[6]Northern Metropolitan Primary Care Directorate, Institut Català de la Salut, Badalona, Spain

[7]Primary Care Services Information Systems, Institut Català de la Salut, Barcelona, Spain

[8]Jordi Gol i Gurina Primary Health Care Research Institute Foundation, Barcelona, Spain

[9]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[10]Interdisciplinary Research Group on ICTs, Barcelona, Spain

[11]Servei Català de la Salut, Barcelona, Spain

[12]Open Evidence Research Group, Universitat Oberta de Catalunya, Barcelona, Spain

[13]Health Department, eHealth Unit, Barcelona, Spain

[14]Pere Virgili Health Park, Primary Care Management Control, Barcelona, Spain

[15]Centre de Recerca en Economia i Salut, Pompeu Fabra University, Barcelona, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, MPH, PhD
Health Promotion in Rural Areas Research Group
Gerència Territorial de la Catalunya Central
Institut Català de la Salut
Carrer Pica d'Estats, 36
Sant Fruitós de Bages, 08272
Spain
Phone: 34 936930040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: eConsulta—that is, asynchronous, two-way teleconsultation in primary care—is one of the most important telemedicine developments in the Catalan public health system, a service that has been heavily boosted by the onset of the COVID-19 pandemic. It is vital to know the characteristics of its users in order to be able to meet their needs and understand the coverage of this service in a context where there is reduced accessibility to the health system.

Objective: This study aims to analyze the profile of the citizens who use the eConsulta tool and the reasons for their use, as well as to gain an understanding of the elements that characterize their decision to use it while distinguishing between those who used it before and those who have used it since the onset of the COVID-19 pandemic.

Methods: A descriptive, observational study based on administrative data was performed. This study differentiates between the COVID-19 pandemic era and the period preceding it, considering the day the state of emergency was declared in Spain (ie,

XSL•FO
RenderX

March 12, 2020) as the cut-off point. It also differentiates between eConsulta users who send messages and those who only receive them.

Results: During the pandemic, the number of unique users of this teleconsultation service had almost tripled, with up to 33.10 visits per 1000 inhabitants per month reported in the first three months. For the two user profiles analyzed, most users since the start of the COVID-19 outbreak were predominantly female, systematically younger, more actively employed, and with less complex pathologies. Furthermore, eConsulta users received more messages proactively from the health professionals. There was also a relative decrease in the number of conversations initiated by higher-income urban users and an increase in conversations initiated by users in rural areas.

Conclusions: The COVID-19 pandemic has helped to generalize the use of telemedicine as a tool to compensate, to some extent, for the decline in face-to-face visits, especially among younger citizens in Catalonia. Telemedicine has made it possible to maintain contact between citizens and the health care system in the context of maximum complexity.

(J Med Internet Res 2021;23(5):e28629) doi:10.2196/28629

KEYWORDS

teleconsultation; primary care; remote consultation; telehealth; COVID-19; e-consultation

Introduction

The eConsulta tool, which has been in operation since 2015, is one of the most important telemedicine developments by the Catalan public health system. It is an asynchronous, two-way teleconsultation tool used by health professionals and citizens that is part of the patient portal of the public health system, a platform that also allows citizens to securely access their personal health information stored in the personal health folder and to carry out certain clinical and administrative procedures. This service was operational only in the context of primary care [1,2] until it was recently expanded to hospital care.

According to data from the Ministry of Health of Catalonia (email, March 3, 2021), from the inception of eConsulta in October 2015 up to February 22, 2021, a total of 15,569 primary care health professionals (out of about 19,000 with potential access) have carried out 4,263,665 e-consultations involving 1,061,995 citizens (out of a total population of 7.7 million. This service is a new model of care relating to health care professionals, more practical for users, and more efficient for the health system, in general [3,4]. In addition, it empowers citizens and promotes teleworking among health care professionals—an essential factor in the context of the current COVID-19 pandemic, which helps improve their work-life balance [5]. In recent years, even before the onset of the pandemic, the use of the eConsulta tool had grown significantly, among both citizens and health care professionals; however, this only represented a very small proportion in relation to the number of face-to-face visits conducted in the Catalan public primary care system [6], a situation very similar to that of other countries [7,8].

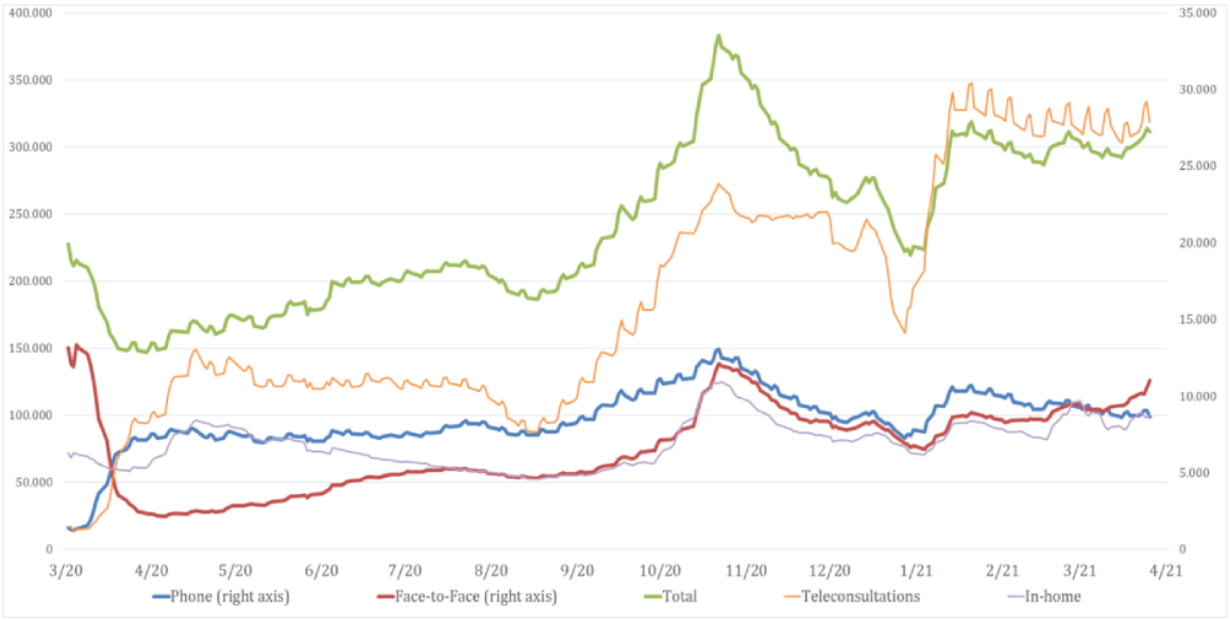
The onset of the COVID-19 pandemic and the initial need to reduce the risk of infection by preventing patients from physically visiting health centers has led to a change in the health care model that has promoted non–face-to-face care. The majority of countries have responded to this need by making significant efforts to implement both synchronous and asynchronous telemedicine approaches [9-12]. The demand for teleconsultation has considerably increased worldwide; for example, in France, the number of teleconsultations increased 50-fold during the initial weeks following the COVID-19 outbreak [13,14]. In Catalonia as well, the outbreak resulted in a very significant increase in the daily number of teleconsultations, although these numbers are well below the number of telephone consultations conducted (Figure 1).

Despite the fact that some studies question the efficiency of teleconsultations between primary care professionals and citizens [7], several studies suggest that the eConsulta tool helps to reduce the number of face-to-face visits [4,6], which may be one of the drivers of its growing use among users. With regard to the health care professionals that use eConsulta, a previous analysis found that doctors who use eConsulta are typically in the 45-64 years age group, score higher than the 80th percentile on the Quality of Care Index, have a high degree of accessibility, are involved in teaching, and work as part of a health team in a high socioeconomic urban setting [15]. With regard to eConsulta users, it is vital to analyze their characteristics in order to provide a service that is appropriate and adjusted to their needs and to also understand who is being covered by this service (and who is not) during the COVID-19 pandemic. Previous studies have shown that women use web-based consultations more frequently than men (64.7% vs 35.3%) and their average age is 39 years [7]. On the contrary, other studies have found that men send email communication for health care more often than women [16]. In Catalonia, however, there is no evidence to determine the profile of a typical citizen or their use of the eConsulta tool.

Thus, the aim of this study is to offer a descriptive analysis of the use of the eConsulta tool before and during the COVID-19 pandemic and the profile of the citizens who use it in order to gain a more comprehensive understanding of the elements that characterize their decision to use the tool and assess who is being covered through this service in a context of low accessibility to the health system.

XSL•FO
RenderX

Figure 1. Primary care daily visits in the Catalonia region of Spain, classified by type with 7-day moving average excluding weekends and public holidays (March 2020 to April 2021). Source: Primary Care Services Information Systems, Institut Català de la Salut, Barcelona, Spain (SISAP) [7].



Methods

This is a descriptive, observational study based on administrative data sourced from the Catalan Health Institute (ICS), the main provider of primary care services in the Catalonia region, serving 74% of the Catalan population. The analysis period was from June 1, 2018, to June 15, 2020. Data of all living patients, as of December 2019 (N=5,844,804), assigned to an ICS primary care team during the study period were analyzed. All these citizens could access and use the eConsulta tool with prior authorization from a health care professional. The analysis differentiates between the COVID-19 pandemic era (ie, from March 12, 2020, the day the state of emergency was declared in Spain, to June 15, 2020) and the preceding period (ie, before March 12, 2020). It also differentiates between eConsulta users who send messages and those who only receive them.

The main study variable is the use of the eConsulta service. “Use” is defined as the period when any messages are sent between a health professional and a citizen, and “nonuse” is defined as the period when no messages have been sent during the study period. The following independent variables were considered in this study: age; gender; socioeconomic level of the center; type of center (rural or urban); adjusted morbidity group (GMA) indicator, a population grouping method that allows the population to be classified into excluding groups according to their multimorbidity [17]; a binary variable identifying the low-income immigrant population; patients with advanced chronic diseases (MACA) indicator [18]; patients with complex chronic diseases (PCC) indicator; and the level of pharmacy coverage. In addition, we assessed the socioeconomic status using the validated MEDEA (Mortality in small Spanish areas and Socioeconomic and Environmental Inequalities) deprivation index, which takes into account the variables of income, occupation, and level of education, among other factors [19]. We categorized this index into quartiles, with the 1st and 4th quartiles representing the least and most deprived areas, respectively. Rural areas were categorized separately and were defined as areas with less than 10,000 inhabitants and a population density lower than 150 inhabitants/km^2.

Continuous variables are presented with mean and SD values, and variables with a nonnormal distribution are presented with median, minimum, and maximum values. Categorical variables are presented with the absolute and relative frequency of each category. For the comparison of two categorical variables, Fisher test and Chi-squared test were used; for the comparison of two numerical variables, the *t* test was used; and in cases where there were more than two variables, the analysis of variance (ANOVA) test was used. A significance level of 5% was set. Data were analyzed using R software (version 3.4.3; R Foundation for Statistical Computing).

Results

Table 1 shows the characterization of citizens who used the service, grouped according to whether they have started a conversation or only received messages, during both the pre–COVID-19 and COVID-19 periods (P<.001). The results show that the profile of the typical user who started a conversation in the pre–COVID-19 era was female (33,096/56,494, 58.6%), with a mean age of 49.84 (17.06) years, a GMA of 2, residing mostly in an urban setting (45,977/56,494, 81.4%). Moreover, the user profile for those who only receive messages was found to be slightly older (mean age: 50.71 [SD 16.16] years), more male (8606/20,104, 42.8%), and more urban (16,492/20,104, 82.0%) compared with active users. When the user profile was analyzed according to the MEDEA deprivation index, we found that since the COVID-19 outbreak, there has been a decrease in the percentage of conversations initiated by higher-income urban citizens (from 4971/20,104, 24.7%, to 19,891/89,102, 22.3%) and an increase in the proportion of users from rural areas (from 10,517/56,494, 18.6%, to 26,958/130,941, 20.6%).

XSL•FO
RenderX

Table 1. Characteristics of citizens who use eConsulta (Source dataset: [20]).

Variable and period	eConsulta users (citizens)		
	Has initiated a conversation (Pre–COVID-19 period: N=56,494; COVID-19 period: N=130,941)	Nonuser (Pre–COVID-19 period: N=5,768,206; COVID-19 period: N=5,624,761)	Only receives messages (Pre–COVID-19 period: N=20,104; COVID-19 period: N=80,102)
Number of messages, mean (SD)			
Pre–COVID-19 period	1.89 (0.39)	N/A[a]	1.06 (0.31)
COVID-19 period	1.70 (0.48)	N/A	1.02 (0.17)
Number of eConsultations, mean (SD)			
Pre–COVID-19 period	4.00 (5.08)	N/A	1.31 (0.75)
COVID-19 period	3.19 (3.05)	N/A	1.83 (1.43)
Age (years), mean (SD)			
Pre–COVID-19 period	49.84 (17.06)	42.67 (23.12)	50.71 (16.16)
COVID-19 period	44.96 (20.04)	42.68 (23.22)	45.02 (15.82)
Gender (female), n (%)			
Pre–COVID-19 period	33,096 (58.58)	2,920,933 (50.64)	11,498 (57.19)
COVID-19 period	75,892 (57.96)	2,839,296 (50.48)	50,339 (56.50)
GMA[b], n (%)			
1			
Pre–COVID-19 period	14,164 (25.07)	2,851,985 (49.44)	4597 (22.87)
COVID-19 period	35,139 (26.84)	2,810,872 (49.97)	24,735 (27.76)
2			
Pre–COVID-19 period	24,033 (42.54)	1,697,448 (29.43)	8519 (42.37)
COVID-19 period	56,896 (43.45)	1,632,369 (29.02)	40,735 (45.72)
3			
Pre–COVID-19 period	13,382 (23.69)	850,893 (14.75)	5317 (26.45)
COVID-19 period	29,522 (22.55)	820,537 (14.59)	19,533 (21.92)
4			
Pre–COVID-19 period	4475 (7.92)	285,988 (4.96)	1548 (7.70)
COVID-19 period	8760 (6.69)	279,505 (4.97)	3746 (4.20)
Immigrant, n (%)			
Pre–COVID-19 period	1369 (2.42)	789,383 (13.69)	642 (3.19)
COVID-19 period	4836 (3.69)	779,631 (13.86)	6927 (7.77)
MACA[c], n (%)			
Pre–COVID-19 period	186 (0.33)	13,167 (0.23)	46 (0.23)
COVID-19 period	348 (0.27)	12,886 (0.23)	165 (0.19)
MEDEA[d], n (%)			
0R[e]			
Pre–COVID-19 period	3222 (5.70)	363,932 (6.31)	1146 (5.70)
COVID-19 period	6001 (4.58)	357,023 (6.35)	5276 (5.92)

XSL•FO
RenderX

Variable and period	eConsulta users (citizens)		
	Has initiated a conversation (Pre–COVID-19 period: N=56,494; COVID-19 period: N=130,941)	Nonuser (Pre–COVID-19 period: N=5,768,206; COVID-19 period: N=5,624,761)	Only receives messages (Pre–COVID-19 period: N=20,104; COVID-19 period: N=80,102)
1R[f]			
Pre–COVID-19 period	2306 (4.08)	331,292 (5.74)	624 (3.10)
COVID-19 period	5965 (4.56)	323,592 (5.75)	4665 (5.24)
2R[g]			
Pre–COVID-19 period	4989 (8.83)	693,257 (12.02)	1842 (9.16)
COVID-19 period	14,992 (11.45)	672,972 (11.96)	12,124 (13.61)
1U[h]			
Pre–COVID-19 period	18,273 (32.35)	1,233,124 (21.38)	4971 (24.73)
COVID-19 period	35,771 (27.32)	1,200,706 (21.35)	19,891 (22.32)
2U[i]			
Pre–COVID-19 period	7168 (12.69)	864,044 (14.98)	2906 (14.45)
COVID-19 period	18,879 (14.42)	842,841 (14.98)	12,398 (13.91)
3U[j]			
Pre–COVID-19 period	13,513 (23.92)	1,190,933 (20.65)	5265 (26.19)
COVID-19 period	31,120 (23.77)	1,159,874 (20.62)	18,717 (21.01)
4U[k]			
Pre–COVID-19 period	7023 (12.43)	1,091,624 (18.92)	3350 (16.66)
COVID-19 period	18,213 (13.91)	1,067,753 (18.98)	16,031 (17.99)
Level of coverage, n (%)			
Active			
Pre–COVID-19 period	44,581 (78.91)	4,262,326 (73.89)	15,268 (75.95)
COVID-19 period	108,917 (83.18)	4,133,267 (73.48)	79,991 (89.77)
Pensioner			
Pre–COVID-19 period	10,890 (19.28)	1,143,779 (19.83)	4311 (21.44)
COVID-19 period	19,233 (14.69)	1,132,752 (20.14)	6995 (7.85%)
PCC[l]			
Pre–COVID-19 period	1695 (3.0)	103,214 (1.79)	459 (2.28)
COVID-19 period	2878 (2.2)	101,411 (1.80)	1079 (1.21)
Rural residents, n (%)			
Pre–COVID-19 period	10,517 (18.62)	1,388,481 (24.07)	3612 (17.97)

XSL•FO
RenderX

Variable and period	eConsulta users (citizens)		
	Has initiated a conversation (Pre–COVID-19 period: N=56,494; COVID-19 period: N=130,941)	Nonuser (Pre–COVID-19 period: N=5,768,206; COVID-19 period: N=5,624,761)	Only receives messages (Pre–COVID-19 period: N=20,104; COVID-19 period: N=80,102)
COVID-19 period	26,958 (20.59)	1,353,587 (24.06)	22,065 (24.76)

[a]N/A: not applicable.

[b]GMA: adjusted morbidity group.

[c]MACA: patients with advanced chronic diseases.

[d]MEDEA: mortality in small Spanish areas and socioeconomic and environmental inequalities

[e]0R: Rural.

[f]1R: Semirural.

[g]2R: Semiurban.

[h]1U: Urban, very high socioeconomic level.

[i]2U: Urban, high socioeconomic level.

[j]3U: Urban, low socioeconomic level.

[k]4U: Urban, very low socioeconomic level.

[l]PCC: patients with complex chronic diseases.

Since the start of the COVID-19 outbreak, eConsulta users have been predominantly female. However, users during the COVID-19 period were systematically younger than those before the pandemic, for the two user profiles analyzed. An increase was observed in the percentage of eConsulta use (especially with regard to the number of messages received from health care professionals) in those population profiles that are more actively employed and have fewer chronic diseases. Thus, we observed a group of working-age citizens who did not used to go to the doctor and citizens who, when they have had to go, preferred the non–face-to-face channel.

In addition, the analysis of the volume of use of the tool during the pre–COVID-19 and COVID-19 periods shows that the number of unique users almost tripled in the first 3 months of the pandemic (from 76,598 to 220,043, 2.87% increase; Table 2). The number of consultations in the 3 months before and after the pandemic showed a monthly increase from 5.61 to 33.10 visits per 1000 inhabitants.

Of these consultations, the proportion of conversations initiated by citizens had reduced (from 74.95% to 52.48%), whereas those initiated by professionals had almost doubled (from 25.05% to 47.52%). Furthermore, the proportion of messages sent by professionals had substantially increased. Similarly, the proportion of conversations involving a response from the professional had decreased considerably (from 1.79% to 0.47%), and the proportion of conversations involving a response initiated by the citizen (ie, patient) had slightly reduced (from 90.8% to 80.61%), as shown in Table 3.

The number of documents attached to eConsulta messages sent by citizens before and during the COVID-19 pandemic was also analyzed (Table 4 and Figure 2). A clear increase in the number of files sent, especially in terms of medical reports, was observed.

Finally, the variations in overall, face-to-face, and web-based care received by users of different age groups were analyzed. Table 5 shows that the use of telemedicine has somewhat mitigated the decline in face-to-face visits in younger age groups.

Table 2. eConsulta activity before and after the COVID-19 pandemic.

Period of analysis	Users (citizens), n (%) (N= 5,844,804)	Consultations, n
Pre–COVID-19 period (June 1, 2018, to March 12, 2020)	76,598 (1.31)	252,370
COVID-19 period (March 13, 2020, to June 15, 2020)	220,043 (3.76)	580,496

XSL•FO
RenderX

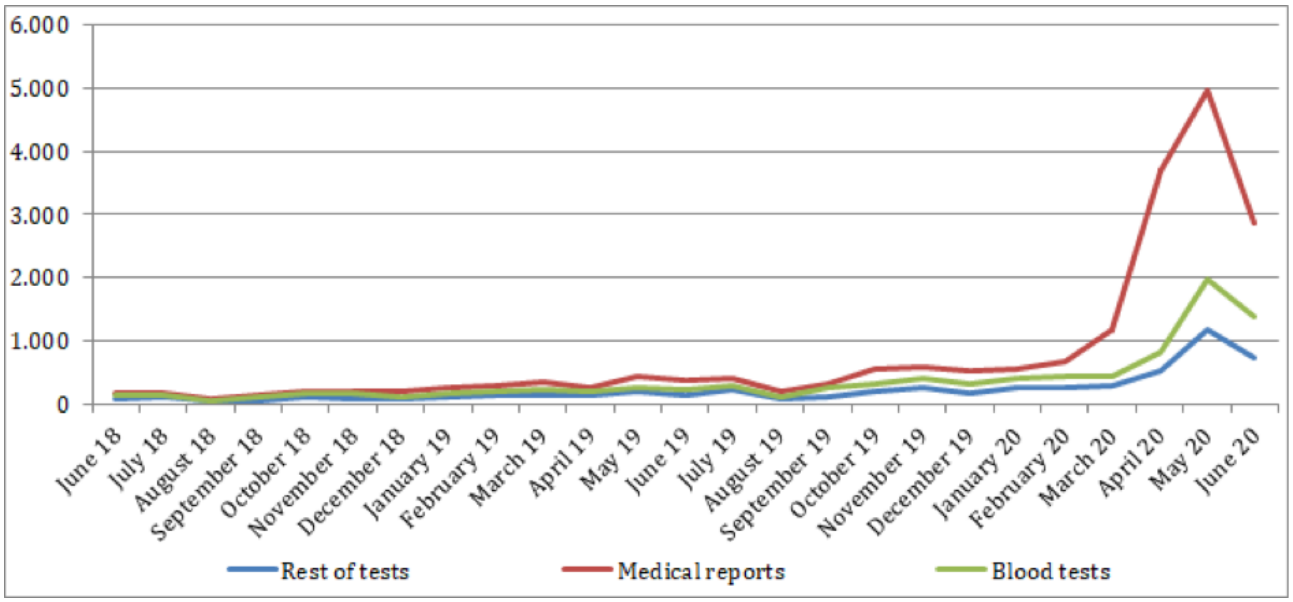
Table 3. Activity by initiator (health care professional or citizen).

Initiator and period	Value, n (%)		
	Number of conversations initiated (Pre–COVID-19 period: N=252,370; COVID-19 period: N=580,496)	Number of conversations with a response	Number of messages sent (Pre–COVID-19 period: N=441,656; COVID-19 period: N=863,867)
Citizens			
Pre–COVID-19 period	189,145 (74.95)	175,534 (92.80)	195,219 (44.20)
COVID-19 period	304,639 (52.48)	245,582 (80.61)	315,656 (36.54)
Health care professionals			
Pre–COVID-19 period	63,225 (25.05)	1,134 (1.79)	246,437 (55.80)
COVID-19 period	275,857 (47.52)	1,286 (0.47)	548,211 (63.46)

Table 4. Average number of monthly tests, classified by type.

Test	Value, mean (SD)		*t* test (*df*)	*P* value
	Pre–COVID-19 period	COVID-19 period		
Medical reports	328.05 (168.02)	3179.50 (1590.69)	3.58 (3.01)	.04
Blood tests	223.62 (107.12)	1151.75 (669.03)	2.77 (3.02)	.07
Other tests	140.57 (64.92)	672.75 (37.01)	2.81 (3.03)	.07

Figure 2. Evolution of the number of documents attached to eConsulta messages, classified by type.



XSL•FO
RenderX

Table 5. Number of face-to-face visits and teleconsultations, classified by age group. Source: Primary Care Services Information Systems, Institut Català de la Salut, Barcelona, Spain (SISAP) [21].

Visit type and age (years)	March to June 2019	March to June 2020	Year-on-year variation (%)
Overall			
<16	28,824	34,722	20.5
16-44	1,756,336	1,786,902	1.7
45-64	2,106,944	2,065,822	–2.0
65-74	1,053,413	706,222	–33.0
>75	1,486,168	1,094,727	–26.3
eConsulta visit , n (%)			
<16	16 (0.06)	550 (1.58)	3337.5
16-44	14,155 (0.81)	173,647 (9.72)	1126.8
45-64	18,072 (0.86)	180,448 (8.73)	898.5
65-74	4,044 (0.38)	16,702 (2.36)	313.0
>75	4,863 (0.33)	16,793 (1.53)	245.3
Face-to-face visit , n (%)			
<16 years	26,188 (90.85)	11,501 (33.12)	–56.1
16-44	1,424,661 (81.12)	427,756 (23.94)	–70.0
45-64	1,614,686 (76.64)	471,869 (22.84)	–70.8
65-74	786,859 (74.7)	200,202 (28.35)	–74.6
>75	1,020,346 (68.66)	316,402 (28.9)	–69.0

Discussion

Principal Findings

The COVID-19 outbreak has led to a significant increase in the use of teleconsultation by both citizens of the Catalonia region and health care professionals. The number of e-consultations per 1000 inhabitants has increased from 5.61, a figure that can be considered low compared to other studies [7], to 33.10 after the start of the COVID-19 pandemic. This increase is explained, first, by the fact that before the onset of the pandemic, patients needed to receive an authorization from their health care professional to be able to carry out e-consultations. After the COVID-19 outbreak, this permission was extended to all citizens. The greater use of teleconsultation is also explained by the increase in the remote provision of care processes that were performed in-person before the pandemic, enabling these citizens to receive remote assistance (ie, sick leave), remote updating of electronic prescription plans for chronic patients, or reactive prescriptions for non–face-to-face visits for acute patients and guidelines for monitoring oral anticoagulant use in patients receiving treatment, among others). The reduction in the average age and the 5-fold increase in “passive” users (ie, users who receive but do not send messages) suggest that eConsulta has been widely used for notifying results or sick leave(s) by allowing users to connect to their personal health folder.

The pandemic has modulated the way eConsulta is used and the typical user profile. There has been an increase in use by both health care professionals, with a clear increase in their initiative in sending messages and documents to citizens, and young patients without chronic diseases. This group comprises citizens, who before the pandemic were infrequent users of the system, but because of the pandemic and their acute pathologies—often related to COVID-19—have had to use the tool in order to contact the health services or receive test results. These processes have made it possible to continue to offer key and prevalent primary care processes in a non–face-to-face and safe and stable manner, by avoiding visits to health centers and contributing decisively to reducing the risk of infection during periods of strict lockdown. It should be noted that eConsulta has not been used for the purpose of COVID-19–related mass messaging by the health system and that it has only been used for the purpose of care continuity for patients.

The results of this study show that it is necessary to understand the type of use of the tool in order to make improvements in its operation and continue working on a model that improves the management of demand for primary care and, in return, its efficiency. To date, the approach based on free-text analysis using machine learning tools seems to be a suitable option to study the evolution over time of the use of the teleconsultation service [3]. This, however, would be a suboptimal solution. The planned evolution of the eConsulta tool for the coming months is precisely the structured stratification of the reasons for consultation, reported by the citizens themselves before initiating it, which will initially allow these messages to be redirected to the professional profiles (doctor, nurse, administrator, dentist, social worker, etc) that can respond in a more agile and appropriate way based on the need expressed by the citizen in their message. In this way, care is decentralized to the different

XSL•FO
RenderX

professional profiles, allowing a more efficient response to the citizen based on the reason for their consultation.

The pandemic caused by the novel coronavirus SARS-CoV-2 has changed the model of health care, and this is especially noticeable in primary care centers, where the most commonly used face-to-face model has been replaced by a mixed model in which telemedicine tools play a very significant role. Although it is clear that the pandemic has led to a reduction in the diagnosis of many diseases [22] and the control of chronic diseases [23], it is necessary to assess the effects of these changes in care (forced by the circumstances) on the health of the supported population in order to continue to guarantee quality health care. We need to emerge from this crisis with a clearer vision of how to continue to deploy telemedicine to obtain its benefits by avoiding or minimizing its drawbacks [12] and inequalities regarding access for the most vulnerable groups [24].

The acceleration of digital transformation processes in health centers has ensured the continuity of care of many basic care processes for the population during lockdown [11]. This new model of care is changing the way we interact with the health care system and the patient profile used by each channel. The results of this study show that non–face-to-face communications from the Catalan primary health care system are being used predominantly in favor of the low-risk and younger population and, therefore, preserving face-to-face and home visits for the most complex and older populations. This adaptation and flexibility of the health system's response, based on the different needs and types of patients, is beneficial and demonstrates the resilience that telemedicine tools have provided to the health system during the COVID-19 pandemic.

This study presents some limitations, as it does not evaluate the relationship between unanswered e-consultations and the type of visit, either in person or by telephone for the same user. It could be possible that the objective of many citizens who have sent an e-consultation was actually to receive a face-to-face or telephone visit in which case the e-consultation would not have replaced these other types of visits.

Conclusions

In the context of reduced face-to-face accessibility to the health system, this study has highlighted a change in the profile of the Catalan citizen using the eConsulta telemedicine tool. Since the start of the COVID-19 pandemic, this patient or user profile is similar to that of the average citizen: actively employed, with less complex pathologies, and who receives more messages proactively from health care professionals through this tool. The COVID-19 pandemic has helped to socialize the use of telemedicine, and as a result, it has helped to mitigate, to some extent, the decline in face-to-face visits in younger age groups.

Conflicts of Interest

None declared.

References

1. Fernández OS, Domínguez CO, Alés XB. Acceso de los pacientes a su historia clínica electrónica: ventajas e inconvenientes para pacientes y profesionales. Article in Spanish. FMC 2017 Oct;24(8):425-427. [doi: 10.1016/j.fmc.2017.04.003]
2. La Meva Salut (Personal Health Folder). Webpage in Catalan. URL: https://lamevasalut.gencat.cat/ [accessed 2021-01-21]
3. López Seguí F, Ander Egg Aguilar R, de Maeztu G, García-Altés A, García Cuyàs F, Walsh S, et al. Teleconsultations between patients and healthcare professionals in primary care in Catalonia: The evaluation of text classification algorithms using supervised machine learning. Int J Environ Res Public Health 2020 Feb 09;17(3):1093 [FREE Full text] [doi: 10.3390/ijerph17031093] [Medline: 32050435]
4. López Seguí F, Walsh S, Solans O, Adroher Mas C, Ferraro G, García-Altés A, et al. Teleconsultation between patients and health care professionals in the Catalan primary care service: Message annotation analysis in a retrospective cross-sectional study. J Med Internet Res 2020 Sep 17;22(9):e19149 [FREE Full text] [doi: 10.2196/19149] [Medline: 32687477]
5. Vidal-Alaball J, López Seguí F. It's here to stay: benefits and drawbacks of eConsultation. Article in Spanish. Atención Primaria Práctica 2020 Jan;2(1-2):100043. [doi: 10.1016/j.appr.2020.100043]
6. López Seguí F, Vidal-Alaball J, Sagarra Castro M, García-Altés A, García Cuyàs F. General practitioners' perceptions of whether teleconsultations reduce the number of face-to-face visits in the Catalan public primary care system: Retrospective cross-sectional study. J Med Internet Res 2020 Mar 16;22(3):e14478 [FREE Full text] [doi: 10.2196/14478] [Medline: 32175914]
7. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
8. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. Health Serv Deliv Res 2018 Jun;6(20). [doi: 10.3310/hsdr06200] [Medline: 29889485]
9. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. Aten Primaria 2020;52(6):418-422 [FREE Full text] [doi: 10.1016/j.aprim.2020.04.003] [Medline: 32402477]

XSL•FO
RenderX

10. Lee I, Kovarik C, Tejasvi T, Pizarro M, Lipoff JB. Telehealth: Helping your patients and practice survive and thrive during the COVID-19 crisis with rapid quality implementation. J Am Acad Dermatol 2020 May;82(5):1213-1214 [FREE Full text] [doi: 10.1016/j.jaad.2020.03.052] [Medline: 32229278]
11. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: Digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
12. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
13. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
14. L'Assurance Maladie. Webpage in French. URL: https://www.ameli.fr/espace-presse/communiques-et-dossiers-de-presse/les-derniers-communiques-de-la-caisse-nationale/detail-d-un-communique/3881.php [accessed 2021-05-17]
15. Fernández OS, Seguí FL, Vidal-Alaball J, Bonet Simo JM, Vian OH, Cabo PR, et al. Primary care doctor characteristics that determine the use of releconsultations in the Catalan public health system: Retrospective descriptive cross-sectional study. JMIR Med Inform 2020 Jan 31;8(1):e16484 [FREE Full text] [doi: 10.2196/16484] [Medline: 32012061]
16. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015 Mar 06;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
17. Monterde D, Vela E, Clèries M, grupo colaborativo GMA. Adjusted morbidity groups: A new multiple morbidity measurement of use in Primary Care. Article in Spanish. Aten Primaria 2016 Dec;48(10):674-682 [FREE Full text] [doi: 10.1016/j.aprim.2016.06.003] [Medline: 27495004]
18. Gómez-Batiste X, Martínez-Muñoz M, Blay C, Espinosa J, Contel JC, Ledesma A. Identifying needs and improving palliative care of chronically ill patients: a community-oriented, population-based, public-health approach. Curr Opin Support Palliat Care 2012 Sep;6(3):371-378. [doi: 10.1097/SPC.0b013e328356aaed] [Medline: 22801465]
19. Garcia-Gil M, Elorza J, Banque M, Comas-Cufí M, Blanch J, Ramos R, et al. Linking of primary care records to census data to study the association between socioeconomic status and cancer incidence in Southern Europe: a nation-wide ecological study. PLoS One 2014;9(10):e109706 [FREE Full text] [doi: 10.1371/journal.pone.0109706] [Medline: 25329578]
20. Vidal-Alaball J. Dataset: Characteristics of citizens who use eConsulta, V1. Mendeley Data. 2021 May 20. URL: http://dx.doi.org/10.17632/ft739dy7jc.1 [accessed 2021-05-20]
21. SISAP (Primary Care Services Information Systems). Webpage in Catalan. URL: https://si9sapics.wordpress.com/about/ [accessed 2021-05-20]
22. Colivicchi F, Di Fusco SA, Magnanti M, Cipriani M, Imperoli G. The impact of the coronavirus disease-2019 Pandemic and Italian lockdown measures on clinical presentation and management of acute heart failure. J Card Fail 2020 Jun;26(6):464-465 [FREE Full text] [doi: 10.1016/j.cardfail.2020.05.007] [Medline: 32417376]
23. Coma E, Mora N, Méndez L, Benítez M, Hermosilla E, Fàbregas M, et al. Primary care in the time of COVID-19: monitoring the effect of the pandemic and the lockdown measures on 34 quality of care indicators calculated for 288 primary care practices covering about 6 million people in Catalonia. BMC Fam Pract 2020 Oct 10;21(1):208 [FREE Full text] [doi: 10.1186/s12875-020-01278-8] [Medline: 33038926]
24. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic consultation in primary care between providers and patients: Systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]

Abbreviations

ANOVA: analysis of variance
ICS: Catalan Health Institute
GMA: adjusted morbidity group
MACA: patients with advanced chronic diseases
MEDEA: Mortality in small Spanish areas and Socioeconomic and Environmental Inequalities
PCC: patients with complex chronic diseases

XSL•FO
RenderX

Edited by C Basch; submitted 09.03.21; peer-reviewed by A Roy, MA Bahrami, J Marconato, T Capote; comments to author 14.04.21; revised version received 29.04.21; accepted 06.05.21; published 27.05.21.

Please cite as:
Solans O, Vidal-Alaball J, Roig Cabo P, Mora N, Coma E, Bonet Simó JM, Hermosilla Pérez E, Saigí-Rubió F, Olmos Domínguez C, Piera-Jiménez J, Abizanda González M, López Seguí F
Characteristics of Citizens and Their Use of Teleconsultations in Primary Care in the Catalan Public Health System Before and During the COVID-19 Pandemic: Retrospective Descriptive Cross-sectional Study
J Med Internet Res 2021;23(5):e28629
URL: https://www.jmir.org/2021/5/e28629
doi:10.2196/28629
PMID:33970867

©Oscar Solans, Josep Vidal-Alaball, Pasqual Roig Cabo, Núria Mora, Ermengol Coma, Josep Maria Bonet Simó, Eduardo Hermosilla Pérez, Francesc Saigí-Rubió, Carmen Olmos Domínguez, Jordi Piera-Jiménez, Mercè Abizanda González, Francesc López Seguí. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

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

Patricia A Areán[1,2*], PhD; Abhishek Pratap[3,4*], PhD; Honor Hsin[5], MD, PhD; Tierney K Huppert[1,6], BA; Karin E Hendricks[1,6,7], MA; Patrick J Heagerty[8], PhD; Trevor Cohen[3], MBChB, PhD; Courtney Bagge[9,10], PhD; Katherine Anne Comtois[1,6], MPH, PhD

[1]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States
[2]ALACRITY Center, University of Washington, Seattle, WA, United States
[3]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States
[4]Sage Bionetworks, Seattle, WA, United States
[5]Kaiser Permanente, Northern California, CA, United States
[6]Center for Suicide Prevention and Research, University of Washington, Seattle, WA, United States
[7]University of South Alabama, Mobile, AL, United States
[8]School of Public Health, University of Washington, Seattle, WA, United States
[9]Department of Psychiatry, University of Michigan Medical Center, Ann Arbor, MI, United States
[10]VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States
[*]these authors contributed equally

Corresponding Author:
Abhishek Pratap, PhD
Department of Biomedical Informatics and Medical Education
University of Washington
850 Republican St
Seattle, WA,
United States
Phone: 1 206 543 2259
Email: apratap@uw.edu

Abstract

Background: Despite decades of research to better understand suicide risk and to develop detection and prevention methods, suicide is still one of the leading causes of death globally. While large-scale studies using real-world evidence from electronic health records can identify who is at risk, they have not been successful at pinpointing when someone is at risk. Personalized social media and online search history data, by contrast, could provide an ongoing real-world datastream revealing internal thoughts and personal states of mind.

Objective: We conducted this study to determine the feasibility and acceptability of using personalized online information-seeking behavior in the identification of risk for suicide attempts.

Methods: This was a cohort survey study to assess attitudes of participants with a prior suicide attempt about using web search data for suicide prevention purposes, dates of lifetime suicide attempts, and an optional one-time download of their past web searches on Google. The study was conducted at the University of Washington School of Medicine Psychiatry Research Offices. The main outcomes were participants' opinions on internet search data for suicide prediction and intervention and any potential change in online information-seeking behavior proximal to a suicide attempt. Individualized nonparametric association analysis was used to assess the magnitude of difference in web search data features derived from time periods proximal (7, 15, 30, and 60 days) to the suicide attempts versus the typical (baseline) search behavior of participants.

Results: A total of 62 participants who had attempted suicide in the past agreed to participate in the study. Internet search activity varied from person to person (median 2-24 searches per day). Changes in online search behavior proximal to suicide

XSL•FO
RenderX

attempts were evident up to 60 days before attempt. For a subset of attempts (7/30, 23%) search features showed associations from 2 months to a week before the attempt. The top 3 search constructs associated with attempts were online searching patterns (9/30 attempts, 30%), semantic relatedness of search queries to suicide methods (7/30 attempts, 23%), and anger (7/30 attempts, 23%). Participants (40/59, 68%) indicated that use of this personalized web search data for prevention purposes was acceptable with noninvasive potential interventions such as connection to a real person (eg, friend, family member, or counselor); however, concerns were raised about detection accuracy, privacy, and the potential for overly invasive intervention.

Conclusions: Changes in online search behavior may be a useful and acceptable means of detecting suicide risk. Personalized analysis of online information-seeking behavior showed notable changes in search behavior and search terms that are tied to early warning signs of suicide and are evident 2 months to 7 days before a suicide attempt.

(J Med Internet Res 2021;23(5):e27918) doi:10.2196/27918

KEYWORDS

real-world data; web searches; suicide risk factors; suicide detection; suicide; eHealth; internet; website; search history; risk; EHR; social media; behavior; mental health; personalized; online seeking behavior

Introduction

Worldwide, suicide is the 18th leading cause of death, resulting in nearly 800,000 lives lost annually [1]. Suicide is the 10th leading cause of death in the US, with nearly 48,000 Americans dying by suicide in 2017 and 1,400,000 suicide attempts in the same year, costing the US approximately 69 billion dollars in 2015 [2,3]. Despite the high societal and personal costs of suicide and decades of research into suicide prevention, we still cannot accurately identify who may be at risk for death by suicide or when the risk is likely to be greatest [4,5].

One of the most important challenges facing suicide prevention researchers—as well as clinical providers—is to identify warning signs for suicidal behavior [6,7], which indicate when a specific individual is at heightened suicide risk in the near term (ie, within minutes, hours, or days) [6,8]. While warning signs developed by expert consensus [6] have been widely disseminated [9-11], research on the timeliness and utility of proposed warning signs has been sparse, and some proposed warning signs have recently been found to not predict imminent suicidal behavior [12]. This is partly because warning signs such as short-term increase in alcohol use [13,14], acute negative interpersonal life events [15], intensity of affective [8], cognitive responses, suicide-related communications, and prepping one's personal affairs [12] may not be static variables (ie, they fluctuate over minutes, hours, or days) making traditional clinical risk assessment at a single health care contact imprecise [6]. Additionally, only one-third of mental health providers use suicide screens routinely [16]. An added complication is that, currently, known methods rely upon having a vulnerable individual self-disclose to and interact with systems (eg, health care, school) or individuals (eg, family, friends) in order to identify warning signs of suicidal behavior. Many people at high risk for suicide do not seek professional help because of lack of time, preference for other sources of help, or fear regarding how they will be treated in the health care system (eg, stigma and the potential trauma associated with in-patient hospitalization) [17-22]. For these reasons, methods that can identify proximal risk factors for suicide that do not depend on health care disclosure and can capture the fluctuating and individualized nature of suicide risk are needed [23].

Machine learning and natural language processing methods have recently been applied to social media data as a means of identifying suicide risk [24,25]. Social media provide a continuous stream of information about individuals' daily lives that may be useful in capturing the dynamic nature of suicide risk, and some studies have found these data can be used to infer a person's mental health as well as ongoing risk for suicide [23,26-29]. A disadvantage to using social media is that the age range of users is still very young [30,31]. Because of recent mishandling of social media data, people are also staying away from public social media platforms [32]. By contrast, 77% of the US population [33] seek information online through internet search engines. The information gleaned here is less likely to be biased, for example, by the desire to project a positive persona on social media [34], with search content focused on information gathering about personal concerns [35,36]. The use of internet search data may be an effective, private, immediate method of proximal suicide risk detection for individuals regardless of their contact with systems of care or self-reported disclosures to other persons.

The use of individualized web searches for proximal risk assessment of suicide attempts is also aligned with the Fluid Vulnerability Theory [29,37-39]. Fluid Vulnerability Theory proposes that suicide risk is a function of ever-changing interactions among multiple risk and protective factors, thus "an individual's vulnerability to suicide is variable but nonetheless identifiable and quantifiable [37]." Some risk and protective factors are static or relatively stable (eg, gender, race, genetics, trauma, dispositional optimism) whereas others are state-based and dynamic (eg, mood, life stressors, insomnia, social support). While static risk factors are known to identify who is at risk, using personalized web searches as a source of real-time real-world behavior may potentially help uncover complex interplay of real-world risk factors associated with when someone is at greatest suicide risk.

However, one limitation to using search data for suicide risk prediction is the concern about privacy and the use of data not intended for public consumption. The intent of social media is to share life events and information publicly whereas search queries are often meant for personal, nonpublic use, and thus privacy concerns about using these data are important to understand [40]. Before embarking on a scaled exploration of

XSL•FO
RenderX

these data for risk detection purposes, it is crucial to gain insight from people with lived experience of suicide about their comfort with their search data being used for suicide detection and prevention.

The purpose of this study is to examine the feasibility of using data from internet searches to identify suicide risk. The first objective of this study was to determine whether internet search behavior (frequency of search queries and queries categorized to known warning signs) were evident within 60 days of a documented suicide attempt. The second objective of this study was to determine how comfortable individuals with lived experience of suicide are with the use of internet search data for early identification.

Methods

Data Collection

Recruitment and Eligibility

Participants with a prior confirmed suicide attempt were recruited from an ongoing randomized clinical trial [41]. Inclusion criteria were (1) inpatient or emergency service admission, (2) lifetime suicide attempt and current hospital admission for suicidality or current hospital admission for a suicide attempt, and (3) consenting to study procedures. This study was approved by the institutional review board at the University of Washington. Data collection occurred from November 2017 to October 2019, and data analysis was conducted from October 2019 to August 2020.

Procedures

Participants were asked to complete a 30-minute semistructured interview about their concerns and suggestions for using internet search data as a means of preventing suicide. They were offered the option to provide a one-time confidential data download of their online Google search history. Participants who opted to participate in the study were reimbursed US $30 regardless of whether or not they agreed to share their Google search history.

Internet Search Data Collection

Google Takeout is a web-based interface developed by Google that allows users of Google apps to download their data into an exportable file. Using prior work [42], we created a web app, called gTAP [43], to allow participants to download their data without sharing personal Google account credentials. Only past Google search history was collected (Multimedia Appendix 1).

The Suicide Attempt and Self-Injury Count, a brief version of the Suicide Attempt Self-Injury Interview [44], has been widely used to determine suicide attempts in clinical trials [45-49] and was used to identify dates of suicide events and categorize events into suicide attempts and nonsuicidal acts.

Participant Survey

The interview was developed by the research team and focused on the acceptability of using internet information to detect the risk of suicide and to prevent suicide (Multimedia Appendix 2). Participants were asked to respond to the following question: “Technology companies use algorithms to predict who is at risk for suicide. Were you aware of this? Do you have any concerns or fears about how this information is collected, stored, and shared? How would you feel if they used your personal search data and/or what you have posted on social media to take action to prevent you and others in a similar situation from suicide? What do you see as the pros and cons?”

Statistical Analysis

Search Data Featurization

Participants’ web searches were used to generate behavioral (online information seeking pattern) and semantic (meaning of search content) features. For behavioral features, we generated daily summary of participants' search history such as the average number of searches per day and the time of day when searches were conducted. For semantic features, we applied distributional models of semantics [50] to derive vector representations of participants’ queries, such that queries and words relating to similar concepts would be proximal in the resulting vector space. To do so, we used semantic vectors [51-53] and publicly available pretrained word embeddings [54] (Multimedia Appendix 3 and Multimedia Appendix 4). In order to map the user search queries to 9 empirically supported warning signs [12] and suicide-method preparation (which has not been previously examined using the present approach), we developed a set of cue terms. An iterative process of cue term definition, expert review of proximal queries, and refinement of the cue term set was used to generate the final set. We calculated a proximity score between the participants' search query (web search) and vector subspace derived from the cue terms representing a warning sign, using the Gram-Schmidt orthonormalization [55] procedure to ensure mutual orthogonality. The proximity score was estimated as the length of the projection of the query into this subspace, following the quantum disjunction method [56]. We used a conservative z score threshold ≥3.5 (corresponding P=.0002) to indicate a meaningful construct-to-query association relative to all search queries for each participant. The use of a threshold of this nature is required on account of the statistical properties of high-dimensional space—all vectors in this space have measurable similarity, but only proximal neighbors in the space indicate meaningful similarity [55,57,58], which should be well above what would be anticipated by chance on account of the high probability of randomly instantiated vectors being mutually close-to-orthogonal [59,60]. Furthermore, this threshold was supported empirically by inspection of the relationships between queries that fell above and the themes of interest in a subset of the data (Multimedia Appendix 4).

Association Analysis

Because of a small sample size and high level of heterogeneity in web search data across participants, we used an individualized analytical approach (Multimedia Appendix 4) to assess the magnitude of difference between web search data proximal (7, 15, 30, and 60 days) to the suicide attempts versus the typical (baseline) search behavior of participants. In order to show the difference between the observed search feature score prior to the time of a suicide attempt and a typical baseline of the search feature for an individual, we computed 2 versions of a standardized value for the search score at the time of the event. First, we presented the z score defined as the search score at the

XSL•FO
RenderX

observed event time (based on a chosen time window), standardized using the mean and variance associated with a patient-specific reference distribution characterizing the full range of observed search scores for an individual. This highlighted how far the observed score at the event time was from typical scores. Second, given that the reference search score distributions may be highly skewed or multimodal, we also computed the empirical reference percentile for each search feature for the event time (suicide attempt). Specifically, we used an empirical reference distribution per search feature to compute the percent of typical individual-specific search feature values that fell either above or below the corresponding search feature value for an observed attempt. A percentile calculation is a form of nonparametric standardization and is termed a *placement value* for classification problems [61]. To show either extremely low or high values, we calculated the symmetrized placement value defined as the minimum of the upper or lower tail probability for the observed attempt score. The participant-specific reference distributions were generated using a nonparametric Monte Carlo simulation [62]. All statistical analysis was performed using the R [63] statistical programming language.

Survey Data Analysis

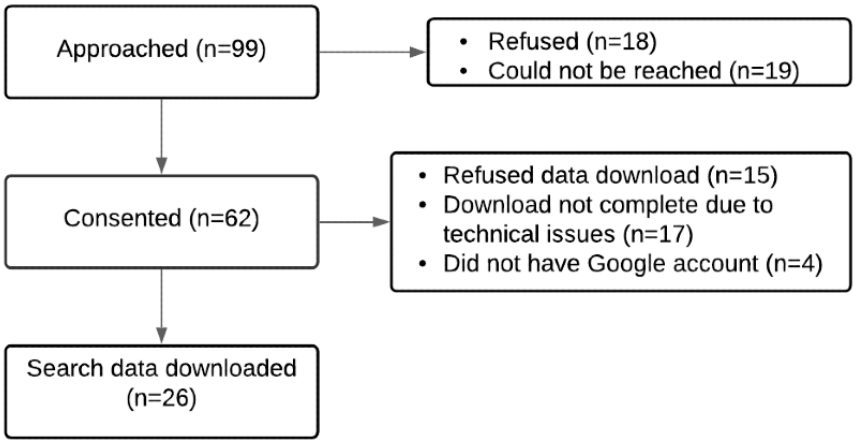
Participants' responses to structured survey questions were summarized using summary statistics. Semistructured responses were transcribed verbatim during participant interviews and anonymized prior to analysis. We used a mixed methods approach, combining quantitative and qualitative data with the function of expansion [64], which allowed inductive qualitative data to provide the *why* to questions uncovered by the quantitative data. Qualitative data were analyzed from a constructivist perspective using thematic analysis after all interviews were completed [65]. Two coders (a psychologist and a psychiatrist) independently extracted initial themes from survey responses. Themes were reviewed by a third coder, a clinical research assistant who had performed all participant interviews. Data were iteratively reviewed (open coded) and collapsed to mutually exclusive themes (axial coding) until saturation was achieved (ie, when no new themes emerged) [66]. Triangulation [67] of quantitative and qualitative data allowed for convergence of themes and a more comprehensive understanding of willingness to use search data for prevention. Illustrative quotes and themes are provided for a qualitative data audit trail. This study was conducted in accordance with SRQR (Standards for Reporting Qualitative Research) [68].

Results

Sample Characteristics

Of 150 individuals, 99 were eligible to participate in this study and were approached, and 62 consented to participate in the qualitative interview. Of the 62 who consented, 26 (42%) were able to provide web search data. Reasons for not providing data were technical issues in downloading search data (17/62, 27%), unwillingness to share Google searches (15/62, 24%), and not having a Google account (4/62, 6%) (Figure 1). The cohort that consented to participate in the qualitative interview consisted of predominantly White individuals (43/62, 69%), and 53.2% were male (33/62). The mean age of the cohort was 34.9 years with 79% (49/62) having at least some college education. No significant differences in demographics were observed between the participants who completed qualitative interviews (n=62) and the final subset (n=26) from whom the Google search data was obtained for age (P=.18), gender (P=.58), race (P=.83), marital status (P=.94), education (P=.96), and income (P=.98) (Table 1). A total of 71 lifetime suicide attempts were reported by the full sample (n=62). The precision of the estimated attempt date varied from the exact attempt date (33/71, 46%) to within 2 weeks (11/71, 15%). To align the suicide attempt period to a proximal web search window, we only used the attempt dates that were rated as accurate within 2 weeks (n=44). Of these, retrospective web search data were available for 30 attempts.

Figure 1. Study CONSORT flow diagram.



XSL•FO
RenderX

Table 1. Demographic characteristics.

Characteristics	Approached (n=99)	Consented (n=62)	Search data downloaded (n=26)	*P* value
Age (years) at enrollment, mean (SD)	33.10 (12.45)	34.94 (13.15)	29.62 (9.15)	.18
Gender, n (%)				.58
Male	50 (50.5)	33 (53.2)	15 (57.7)	
Female	38 (38.4)	21 (33.9)	5 (19.2)	
Other	5 (5.1)	4 (6.5)	3 (11.5)	
Transgender	6 (6.1)	4 (6.5)	3 (11.5)	
Race, n (%)				.83
White	66 (66.7)	43 (69.4)	21 (80.8)	
Mixed	20 (20.2)	14 (22.6)	3 (11.5)	
Asian	7 (7.1)	2 (3.2)	2 (7.7)	
Black or African American	4 (4.0)	3 (4.8)	0 (0.0)	
American Indian or Alaska Native	1 (1.0)	0 (0.0)	0 (0.0)	
Native Hawaiian or Other Pacific Islander	1 (1.0)	0 (0.0)	0 (0.0)	
Marital status, n (%)				.94
Single/never married	72 (72.7)	42 (67.7)	18 (69.2)	
Divorced	12 (12.1)	10 (16.1)	3 (11.5)	
Married	9 (9.1)	4 (6.5)	2 (7.7)	
Separated	5 (5.1)	5 (8.1)	3 (11.5)	
Widowed	1 (1.0)	1 (1.6)	0 (0.0)	
Education, n (%)				.96
Some college, associate's degree, or technical training	53 (53.5)	33 (53.2)	16 (61.5)	
Bachelor's or graduate degree	22 (22.2)	16 (25.8)	4 (15.4)	
High school graduate or GED	15 (15.2)	9 (14.5)	5 (19.2)	
Some high school	7 (7.1)	3 (4.8)	1 (3.8)	
Other	2 (2.0)	1 (1.6)	0 (0.0)	
Income[a], n (%)				.98
Less than $5000	9 (10.8)	5 (9.4)	3 (14.3)	
$5000-9999	11 (13.3)	8 (15.1)	2 (9.5)	
$10,000-24,999	23 (27.7)	12 (22.6)	4 (19.0)	
$25,000-49,999	21 (25.3)	16 (30.2)	7 (33.3)	
More than $50,000	15 (18.1)	9 (17.0)	4 (19.0)	
None	4 (4.8)	3 (5.7)	1 (4.8)	

[a]Data were missing for n=16, n=9, and n=5 individuals for approached, consented, and data downloaded, respectively.

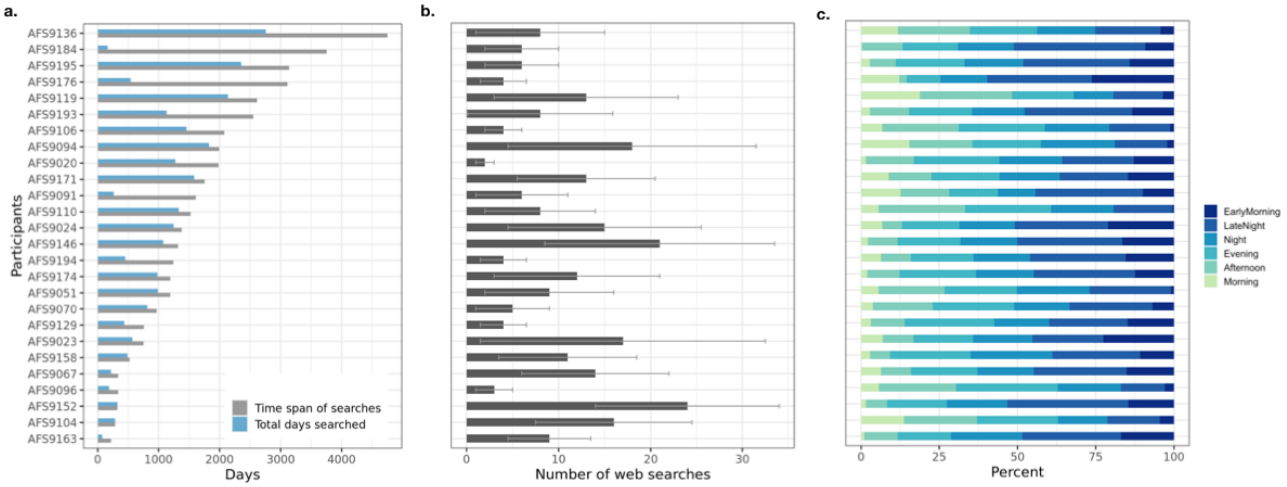
Search Data

Summary

In total, search history data for 24,397 days were collected with 349,922 individual search queries from 26 study participants. The median time span of the search history data across participants was 1348 days (range 220-4752 days); however, the actual number of days when participants searched online was much lower than the data collection period and varied widely (median 898.5 days, range 75-2759 days) (Figure 2a). Analysis generated 11 high-level search constructs related to participants' online information-seeking behavior (ie, search behavior features) and search content proximity scores for 10 suicide warning signs (ie, semantic features) (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Search data characteristics across participants: (a) span (in days) of search data collected from participants (in grey) and the number of days (blue) on which participants made at least one search. (b) Median daily number of web searches performed by the participants. The error bars indicate the 25th and 75th percentile. (c) Proportions of participants' web searches stratified by time of day.



Search Behavior Features

Our analysis revealed idiosyncratic online information seeking behavior across individuals; the number of searches per day varied between 2 searches and 24 searches (median 8.5 searches). The time of day when participants conducted online searches also varied (morning: 0%-4.35%; late night: 0%-37.5%; Figure 2b,c).

Semantic Features

Of the semantic search content that we mapped to known suicide warning signs, we identified a small proportion of search queries (median 1.2%, range 0.06%-21.47%) with a proximity *z* score ≥3.5 that mapped onto 1 or more of the 10 warning signs. At times, queries meeting this threshold were observed in close proximity to a suicide attempt (within days to weeks). Table 2 provides a representative selection of highly ranked queries (based on *z* scores) for each warning sign, revealing content suggestive of premeditation, preparation, substance use, stressors, underlying mental state, and help-seeking behavior.

Table 2. Cue term sets developed to represent selected warning signs and a subset of top search queries that map to each of the warning signs.

Warning sign	Cue terms	Retrieved search queries
Alcohol use	whiskey; alcohol; aa; wine; alcoholic; beer	"aa meetings"; "how much beer to get drunk"; "wine hangover vs. hard alcohol"; "alcohol poisoning"; "alcoholics anonymous"[a]
Preparation of personal affairs	will; affairs; suicide+note	"writing a suicide note"; "living will"; "write your will online"
Suicide communication	hotline; help; suicide+communicate	"what does suicide hotline do"; "suicide crisis text line"; "suicide text line"; "emergency room si suicidal ideation"
Suicide methods (preparation)	overdose; gun; lethal	"sleeping pill overdose suicide"; "is ambien lethal"; "where can I get suicide pills"[a]; "where to buy a gun in Seattle"; "cheap guns"[a]
Burdensomeness	burden	"discussing work burdens marriage"[a]
No reason to live	hopeless; live; persist	"I don't want to live anymore"
Anger	hostile; rage; anger	"fits of rage"; "depression and rage"; "serious anger marijuana"
Anxiety	scared; fearful; afraid; anxiety; anxious; jittery;	"ocd anxiety"; "apprehensive"[a]; "social anxiety"; "marijuana for anxiety"; "why do I have so much anxiety"; "phobia of diseases"[a]
Emptiness	numb; hollow; feeling+empty	"I feel so empty"; "I like the feeling of being sad"
Interpersonal problem	conflict; divorce; fight; breakup; loss	"final divorce decree cost"; "infidelity and custody"[a]; "how much child support if spouse loses job"[a]; "divorce"

[a]Found using distributional semantic approaches (ie, queries do not contain any of the manually defined cue terms) illustrating the capacity of distributional semantics approaches to identify related concepts expressed in different terms.

XSL•FO
RenderX

Association Between Search Data Features and Suicide Attempts

On average, 58% of attempts (n=30; range 15/30, 50% to 19/30, 63%) were found to be associated ($-\log_{10}$(placement value)≥2) with at least one search feature in 1 of the 4 proximal time periods (7, 15, 30, and 60 days). Figure 3 shows the summary of individualized association analysis highlighting the specific search constructs proximally associated with suicide attempts. Notably, for 23% attempts (7/30), a prolonged association with proximal search features (from 60 days) was observed, indicating an extended period of potentially high-risk online search behavior. For the majority of attempts, we observed a high degree of variation in search constructs and the proximal time period in which they were associated with attempts. The constructs associated the most attempts (across time windows) were online search patterns (9/30 attempts, 30%), semantic similarity of search queries to suicide methods (7/30, 23%), and anger (7/30, 23%). Figure 4 shows features associated with 4 individual suicide attempts (one per proximal window) where the search behaviors of the participants were found to be markedly different ($-\log_{10}$(placement value)≥2) from their typical (ie, baseline) behavior.

Figure 3. Summary of individualized association analysis for 11 high-level search constructs over 4 suicide attempt–proximal periods: (a) 7, (b) 15 (c) 30, and (d) 60 days.

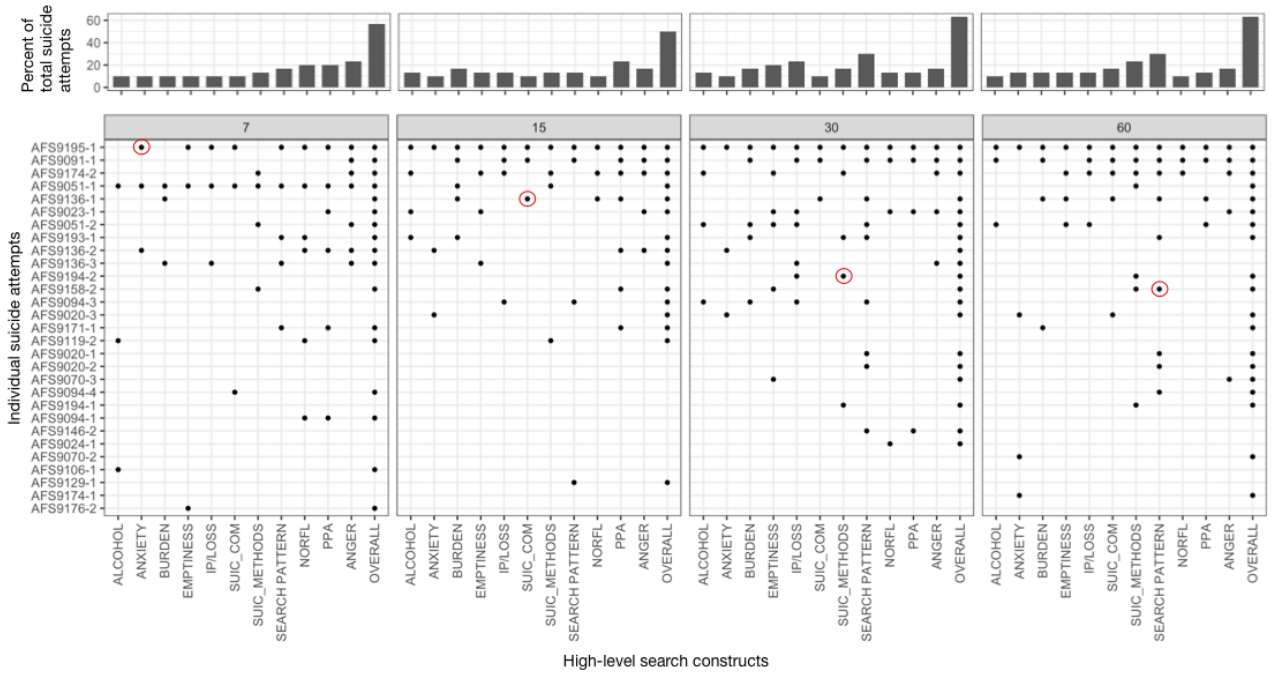
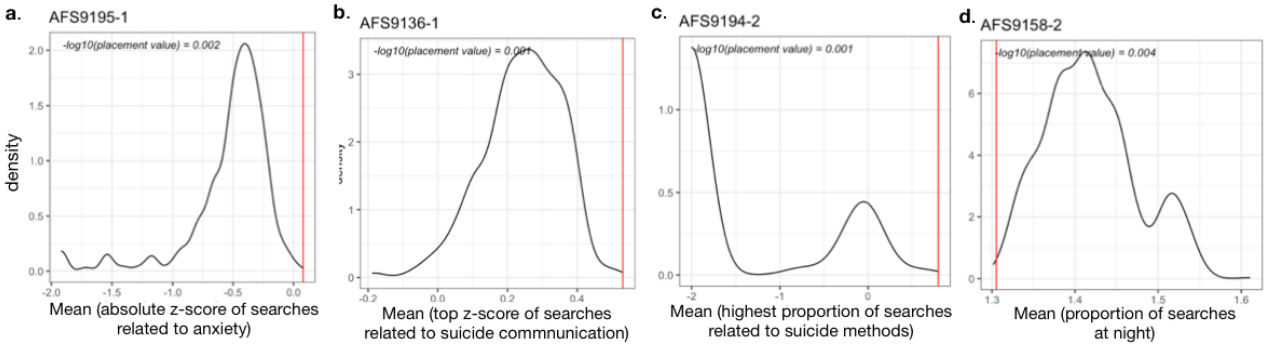


Figure 4. Baseline distributions for 4 example search features (each indicated by a red circle in Figure 3): (a) anxiety – 7-day proximal period; (b) suicide communication – 15-day proximal period; (c) suicide methods – 30-day proximal period; and (d) searches at night – 60-day proximal period. The red line indicates the value of the search feature in the corresponding time period proximal to a suicide attempt.



Perceived Acceptability and Feasibility of Internet Search–Based Prediction of Suicide Risk

Three primary themes were identified regarding the acceptability of using search history for suicide prevention: utility, accuracy, and privacy (Table 3): 68% (40/59) thought using internet search history for suicide detection and prevention could be useful; many participant (34/59, 58%) raised concerns about the accuracy of detection, in particular concerns of false positives and their potential consequences; and 32% of participants (19/59) were concerned about the intrusion of privacy.

XSL•FO
RenderX

Table 3. Illustrative quotes of participant responses to use of internet history in suicide prevention.

Theme	Illustrative quotation	Respondents reporting the theme, n (%)
Useful	"It'd be a good way to help people get resources that they don't otherwise know about."	40/59 (68)
	"I think it's good. I think people would be more open online than how they are in one-to-one in-person situations."	
	"I think it sounds great. They're already using algorithms to optimize search, [so] might as well do it for something good."	
	"Potentially it could be enough to 'break the cycle' of suicidal thoughts."	
Detection accuracy concerns	"No problem with that as long as they did it right. I wouldn't want the SWAT team to show up at my door..."	34/59 (58)
	"Companies don't understand the context of the situation, try to do the right thing, but...it makes people want to shut people out because they overreact."	
	"I don't think I would mind because I know what I'm getting myself into. They should work on the algorithm before implementing it into the general population though."	
	"I like how they use trigger/key words to dump all these resources on you, but I think they need to improve it."	
Privacy concerns	"I'm chronically in private mode, because I don't want Google or tech other companies knowing I'm looking at this. If I'm ever in public, I don't want my search results to be seen by others."	19/59 (32)
	"Feels like I am being spied on."	
	"I would be a little upset about that. Seems like an invasion of privacy."	
	"[As] we become more transparent on the internet, search history or data could be used [with] malicious intentions, planting bombs for the future."	

When presented with potential prevention interventions, participants favored interventions that provided a direct link to either a crisis counselor (35/61, 57%), friends or family (33/60, 55%), peers (30/61, 49%), or to a self-guided meditation video (33/61, 54%) (Multimedia Appendix 5). Interventions that simply provided a hotline number, with suggestions to reach out, or an inspirational video were not as favorable. When asked if they had seen interventions such as Google's links to a suicide prevention hotline, only 53% (33/62) said they saw the link, and of those, only one person acted on it. Participants were allowed to voluntarily opt out of answering any questions in the qualitative survey, and if they did so for a specific question, they were not counted in the denominator of that question's response percentages. Although 62 participants consented to the qualitative interview and provided some form of response in the interview, up to 3 participants opted out of specific questions at various times for declining to answer or feeling too distressed to continue the interview.

Discussion

General

This is one of the first studies to examine and describe the nature of individualized internet search data with an eye toward suicide prevention. We found that while search queries and behavior do change prior to suicide attempts, there is considerable variation between individuals, with some participants searching online more frequently, and others seeking information online sporadically prior to attempts. Additionally, search queries over time are highly individualized, and for some attempts, changes in search behavior and queries related to risk are evident 60 days before the attempt, with a majority evident 2 weeks before the attempt. Search content associated with risk windows also varied, although some content was highly prevalent across time points such as queries expressing anger or suicide methods. Although these findings suggest that the use of internet searches for risk prediction will be complicated due to the intraindividual variation, it may still be possible to develop a personalized temporal risk profile or a digital phenotype [69] linked to suicide-related behaviors. Previous research has found that personalized models lead to more accurate prediction of clinical states [70]. In the present study, individualized risk assessment analysis identified as much as 63% of attempts (19/30) based on changes in search behavior and queries.

We found that participants felt using internet search data to predict and intervene in suicide was potentially helpful, but they also harbored some important reservations. Participants felt that any intervention based on search history or social media algorithms would need to be highly accurate and respect personal privacy. The interventions themselves should be active (link to a friend), rather than passive (suggestion to contact a hotline). Importantly, participants were particularly concerned about the use of emergency services as a means of intervention.

This study represents the first step in understanding the potential utility of online search data for suicide prevention. The next steps will require a study with a much larger sample size due to the intraindividual variation in search signal differences, in addition to interindividual variation in search terms and search behaviors prior to attempts. Expansion of the semantic feature space may also further refine predictive signals. While these results demonstrate that a personalized analytical approach can identify patterns of search behaviors that are evident up to 2 months before an attempt, larger studies are needed to assess potential representational bias and further refine high-risk signatures from online search data.

XSL•FO
RenderX

Limitations

This study is a preliminary cohort study and thus has limitations. First, this was a small sample. Although individualized analysis of web search data indicates the potential benefits for understanding real-world risk factors of suicide and when someone may be at a higher risk, future research should explore the cohort-level predictive ability of data, including optimization of analytical parameters (eg, selection of threshold to indicate a meaningful association between web searches and suicide risk factors). Second, participants in this study were at very high suicide risk with both a lifetime suicide attempt and a recent episode requiring hospitalization. A larger prospective study of people with varying levels of suicide risk, including those without a history of suicidal ideation, is warranted to ensure that the search patterns and terms found here are unique to the imminent risk of suicide. Understanding the perspectives of other individuals on the sharing of web search data and the appropriateness of intervention will also be crucial prior to any deployment of prediction algorithms and related suicide prevention efforts. Finally, we asked participants their perspectives about hypothetical intervention scenarios based on internet search informed risk prediction. It is likely that participants' acceptability of such interventions will differ when they are faced with them during a crisis event.

Conclusion

Although this is a preliminary study, the findings are promising and suggest a potentially useful and timely method for utilizing search data for detecting the risk of suicide. If handled appropriately, this method of risk detection is seen by those with a lived experience of suicide as an acceptable method of detection. Suicide is a serious public health problem, one that has the potential to escalate during these times of health, societal and economic challenges. Methods that can quickly identify and intervene to prevent a suicide event could help prevent and reduce the public health burden of suicide.

Acknowledgments

This work was supported in part by grants from the National Institute of Mental Health (P50 MH115837, R01 MH10230f4, and R33 MH110509).

Authors' Contributions

PAA, AP, and KC designed the study. AP and TC led the search and Google Search data analyses with contributions from PH and CB. KH and TH conducted all study interviews. PAA, HH, and TH contributed to the qualitative analysis of the study and interpretation of the descriptive analyses. PAA, KC, AP, TC, and HH were involved in writing the first draft of the manuscript. PAA obtained study funding. All authors critically revised the manuscript.

Conflicts of Interest

PAA reports consulting with Verily Life Sciences on mental health and technology projects. HH was an employee of Verily Life Sciences prior to study contribution.

Multimedia Appendix 1
A brief schematic overview of the gTAP workflow.
[PDF File (Adobe PDF File), 238 KB - jmir_v23i5e27918_app1.pdf]

Multimedia Appendix 2
Interview Guide for Exploring Technology for Suicide Prevention AFS Sub-Study.
[PDF File (Adobe PDF File), 146 KB - jmir_v23i5e27918_app2.pdf]

Multimedia Appendix 3
Description of Search Data Features.
[PDF File (Adobe PDF File), 93 KB - jmir_v23i5e27918_app3.pdf]

Multimedia Appendix 4
Schematic overview of search data featurization and non-parametric association analysis to compare the difference between a typical "baseline" search behavior to a proximal period before the documented suicide attempt.
[PDF File (Adobe PDF File), 648 KB - jmir_v23i5e27918_app4.pdf]

Multimedia Appendix 5
Survey participants were asked to rate helpfulness and comfortableness with each of the following potential intervention options. Values represent percentages of all participants that responded to each question.
[PDF File (Adobe PDF File), 72 KB - jmir_v23i5e27918_app5.pdf]

XSL•FO
RenderX

References

1. Suicide data. World Health Organization. URL: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ [accessed 2020-09-23]
2. Suicide mortality in the United States, 1999-2017. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/products/databriefs/db330.htm [accessed 2021-02-10]
3. Suicide statistics. American Foundation for Suicide Prevention. URL: https://afsp.org/suicide-statistics [accessed 2021-04-16]
4. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
5. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019 Jun 01;76(6):642-651. [doi: 10.1001/jamapsychiatry.2019.0174] [Medline: 30865249]
6. Rudd M, Berman AL, Joiner TE, Nock MK, Silverman MM, Mandrusiak M, et al. Warning signs for suicide: theory, research, and clinical applications. Suicide Life Threat Behav 2006 Jun;36(3):255-262. [doi: 10.1521/suli.2006.36.3.255] [Medline: 16805653]
7. Glenn CR, Nock MK. Improving the short-term prediction of suicidal behavior. Am J Prev Med 2014 Sep;47(3 Suppl 2):S176-S180 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.004] [Medline: 25145736]
8. Bagge CL, Littlefield AK, Glenn CR. Trajectories of affective response as warning signs for suicide attempts. Clin Psychol Sci 2017 Mar 13;5(2):259-271 [FREE Full text] [doi: 10.1177/2167702616681628]
9. Warning signs of suicide. Suicide Awareness Voices of Suicide. URL: https://save.org/about-suicide/warning-signs-risk-factors-protective-factors/ [accessed 2021-02-12]
10. Risk factors and warning signs. American Foundation for Suicide Prevention. URL: https://afsp.org/risk-factors-and-warning-signs [accessed 2021-04-16]
11. Recognizing suicidal behavior. WebMD. URL: https://www.webmd.com/mental-health/recognizing-suicidal-behavior [accessed 2021-02-12]
12. Bagge C. An examination of a broad list of warning signs for suicide attempts: a multisite investigation. 2019 Presented at: International Summit on Suicide Research; October 27-30; Miami, Florida.
13. Bagge CL, Borges G. Acute substance use as a warning sign for suicide attempts. J Clin Psychiatry 2017 Jun 28;78(06):691-696. [doi: 10.4088/jcp.15m10541]
14. Borges G, Bagge CL, Cherpitel CJ, Conner KR, Orozco R, Rossow I. A meta-analysis of acute use of alcohol and the risk of suicide attempt. Psychol Med 2016 Dec 08;47(5):949-957. [doi: 10.1017/s0033291716002841]
15. Bagge CL, Glenn CR, Lee H. Quantifying the impact of recent negative life events on suicide attempts. J Abnorm Psychol 2013 May;122(2):359-368. [doi: 10.1037/a0030371] [Medline: 23088374]
16. Roush JF, Brown SL, Jahn DR, Mitchell SM, Taylor NJ, Quinnett P, et al. Mental health professionals' suicide risk assessment and management practices. Crisis 2018 Jan;39(1):55-64. [doi: 10.1027/0227-5910/a000478] [Medline: 28914092]
17. Tondo L, Albert MJ, Baldessarini RJ. Suicide rates in relation to health care access in the United States: an ecological study. J Clin Psychiatry 2006 Apr 15;67(4):517-523. [doi: 10.4088/jcp.v67n0402] [Medline: 16669716]
18. Ko J, Frey JJ, Harrington D. Preventing suicide among working-age adults: the correlates of help-seeking behavior. Inquiry 2019 May 26;56:46958019850979 [FREE Full text] [doi: 10.1177/0046958019850979] [Medline: 31130054]
19. Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. JAMA 2005 Oct 26;294(16):2064-2074. [doi: 10.1001/jama.294.16.2064] [Medline: 16249421]
20. Reynders A, Kerkhof AJFM, Molenberghs G, Van Audenhove C. Attitudes and stigma in relation to help-seeking intentions for psychological problems in low and high suicide rate regions. Soc Psychiatry Psychiatr Epidemiol 2014 Feb 30;49(2):231-239 [FREE Full text] [doi: 10.1007/s00127-013-0745-4] [Medline: 23896893]
21. Blais RK, Renshaw KD. Stigma and demographic correlates of help-seeking intentions in returning service members. J Trauma Stress 2013 Feb 18;26(1):77-85. [doi: 10.1002/jts.21772] [Medline: 23335155]
22. Hom MA, Stanley IH, Schneider ME, Joiner TE. A systematic review of help-seeking and mental health service utilization among military service members. Clin Psychol Rev 2017 Apr;53:59-78. [doi: 10.1016/j.cpr.2017.01.008] [Medline: 28214634]
23. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018 Aug 27;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
24. Shen J, Rudzicz F. Detecting anxiety through Reddit. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology -- From Linguistic Signal to Clinical Reality. 2017 Presented at: Fourth Workshop on Computational Linguistics and Clinical Psychology; August; Vancouver p. 58-65. [doi: 10.18653/v1/w17-3107]
25. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Proc Int AAAI Conf Weblogs Soc Media 2017 May;2017:32-41 [FREE Full text] [Medline: 28840079]

XSL•FO
RenderX

26. Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide Life Threat Behav 2017 Feb 03;47(1):112-121. [doi: 10.1111/sltb.12312] [Medline: 27813129]
27. Choudhury MD, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4; Ann Arbor, Michigan.
28. Bryan C, Butner J, Sinclair S, Bryan A. Patterns of change in user-generated social media behavior among service members who die by suicide. Defense Personnel and Security Research Center Seaside United States. 2016. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/1036617.pdf [accessed 2021-04-16]
29. Bryan CJ, Butner JE, Sinclair S, Bryan ABO, Hesse CM, Rose AE. Predictors of emerging suicide death among military personnel on social media networks. Suicide Life Threat Behav 2018 Aug 28;48(4):413-430. [doi: 10.1111/sltb.12370] [Medline: 28752655]
30. Demographics of social media users and adoption in the United States. Pew Research Center. 2018. URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-05-04]
31. Coppersmith DD, Kleiman EM, Glenn CR, Millner AJ, Nock MK. The dynamics of social support among suicide attempters: a smartphone-based daily diary study. Behav Res Ther 2019 Sep;120:103348 [FREE Full text] [doi: 10.1016/j.brat.2018.11.016] [Medline: 30594300]
32. Pratap A, Allred R, Duffy J, Rivera D, Lee HS, Renn BN, et al. Contemporary views of research participant willingness to participate and share digital data in biomedical research. JAMA Netw Open 2019 Nov 01;2(11):e1915717 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15717] [Medline: 31747031]
33. Search engine market share United States Of America. StatCounter Global Stats. URL: http://gs.statcounter.com/search-engine-market-share/all/united-states-of-america [accessed 2019-05-04]
34. Carney M. The pressure to have a perfect social media presence. Massachusetts Daily Collegian. URL: https://dailycollegian.com/2018/10/the-pressure-to-have-a-perfect-social-media-presence/ [accessed 2019-05-04]
35. Adler N, Cattuto C, Kalimeri K, Paolotti D, Tizzoni M, Verhulst S, et al. How search engine data enhance the understanding of determinants of suicide in india and inform prevention: observational study. J Med Internet Res 2019 Jan 04;21(1):e10179 [FREE Full text] [doi: 10.2196/10179] [Medline: 30609976]
36. Asch JM, Asch DA, Klinger EV, Marks J, Sadek N, Merchant RM. Google search histories of patients presenting to an emergency department: an observational study. BMJ Open 2019 Feb 20;9(2):e024791 [FREE Full text] [doi: 10.1136/bmjopen-2018-024791] [Medline: 30787088]
37. Rudd MD. Cognition and suicide: theory, research, and therapy. In: Ellis TE, editor. Fluid Vulnerability Theory: A Cognitive Approach to Understanding the Process of Acute and Chronic Suicide Risk. Washington, DC: American Psychological Association; 2006:355-368.
38. Bryan CJ, Rudd MD. The importance of temporal dynamics in the transition from suicidal thought to behavior. Clin Psychol (New York) 2016 Mar 11;23(1):21-25 [FREE Full text] [doi: 10.1111/cpsp.12135]
39. Bryan CJ, Rudd MD. Nonlinear change processes during psychotherapy characterize patients who have made multiple suicide attempts. Suicide Life Threat Behav 2018 Aug 09;48(4):386-400. [doi: 10.1111/sltb.12361] [Medline: 28597959]
40. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
41. Reducing short term suicide risk after hospitalization. American Foundation for Suicide Prevention. URL: https://afsp.org/grant/reducing-short-term-suicide-risk-after-hospitalization [accessed 2021-04-16]
42. Zaman A, Acharyya R, Kautz H, Silenzio V. Detecting low self-esteem in youths from web search data. 2019 Presented at: International World Wide Web Conference; May 13-17; San Francisco, California. [doi: 10.1145/3308558.3313557]
43. Pratap A. gTap. GitHub. URL: https://github.com/apratap/gTap [accessed 2020-05-29]
44. Linehan MM, Comtois KA, Brown MZ, Heard HL, Wagner A. Suicide Attempt Self-Injury Interview (SASII): development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury. Psychol Assess 2006 Sep;18(3):303-312. [doi: 10.1037/1040-3590.18.3.303] [Medline: 16953733]
45. Comtois KA, Kerbrat AH, DeCou CR, Atkins DC, Majeres JJ, Baker JC, et al. Effect of augmenting standard care for military personnel with brief caring text messages for suicide prevention: a randomized clinical trial. JAMA Psychiatry 2019 May 01;76(5):474-483 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.4530] [Medline: 30758491]
46. Jobes DA, Comtois KA, Gutierrez PM, Brenner LA, Huh D, Chalker SA, et al. A randomized controlled trial of the collaborative assessment and management of suicidality versus enhanced care as usual with suicidal soldiers. Psychiatry 2017 Feb 21;80(4):339-356. [doi: 10.1080/00332747.2017.1354607] [Medline: 29466107]
47. van den Bosch LM, Verheul R, Schippers GM, van den Brink W. Dialectical behavior therapy of borderline patients with and without substance use problems. Addict Behav 2002 Nov;27(6):911-923. [doi: 10.1016/s0306-4603(02)00293-9]
48. Luxton DD, Thomas EK, Chipps J, Relova RM, Brown D, McLay R, et al. Caring letters for suicide prevention: implementation of a multi-site randomized clinical trial in the U.S. military and Veteran Affairs healthcare systems. Contemp Clin Trials 2014 Mar;37(2):252-260. [doi: 10.1016/j.cct.2014.01.007] [Medline: 24473106]

XSL•FO
RenderX

49. Ryberg W, Fosse R, Zahl PH, Brorson I, Møller P, Landrø NI, et al. Collaborative Assessment and Management of Suicidality (CAMS) compared to treatment as usual (TAU) for suicidal patients: study protocol for a randomized controlled trial. Trials 2016 Oct 03;17(1):481 [FREE Full text] [doi: 10.1186/s13063-016-1602-z] [Medline: 27716298]
50. 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]
51. 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; May 31-June 1; Marrakech, Morocco URL: http://www.lrec-conf.org/proceedings/lrec2008/pdf/300_paper.pdf [doi: 10.1057/9781137294678.0479]
52. Widdows D, Cohen T. The Semantic Vectors package: new algorithms and public tools for distributional semantics. In: Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing. 2010 Sep 01 Presented at: IEEE Fourth International Conference on Semantic Computing; September 22-24; Pittsburgh, Pennsylvania p. 9-15. [doi: 10.1109/icsc.2010.94]
53. Semanticvectors. GitHub. URL: https://github.com/semanticvectors/semanticvectors [accessed 2020-06-25]
54. English word vectors. Facebook Open Source. URL: https://fasttext.cc/docs/en/english-vectors.html [accessed 2020-09-16]
55. Mayers DF, Golub GH, van Loan CF. Matrix Computations. Baltimore: Johns Hopkins University Press; Jul 1996:694.
56. Widdows D, Peters S. Word vectors and quantum logic experiments with negation and disjunction. 2003 Presented at: Eighth Meeting on Mathematics of Language; June 20-22; Bloomington, Indiana p. 1-1 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.74.7623&rep=rep1&type=pdf
57. Karlgren J, Holst A, Sahlgren M. Filaments of meaning in word space. In: Macdonald C, Ounis I, Plachouras V, Ruthven I, White RW, editors. Advances in Information Retrieval. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer; 2008:531-538.
58. Rekabsaz N, Lupu M, Hanbury A. Exploration of a threshold for similarity based on uncertainty in word embedding. In: Jose J, editor. Advances in Information Retrieval. Lecture Notes in Computer Science. Cham: Springer; 2017:396-409.
59. Kanerva P. Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. Cogn Comput 2009 Jan 28;1(2):139-159 [FREE Full text] [doi: 10.1007/s12559-009-9009-8]
60. Widdows D, Cohen T. Reasoning with vectors: a continuous model for fast robust inference. Log J IGPL 2015 Oct 19;23(2):141-173. [doi: 10.1093/jigpal/jzu028] [Medline: 26582967]
61. Hanley JA, Hajian-Tilaki KO. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. Acad Radiol 1997 Jan;4(1):49-58. [doi: 10.1016/s1076-6332(97)80161-4]
62. Raychaudhuri S. Introduction to Monte Carlo simulation. : IEEE; 2008 Presented at: 2008 Winter Simulation Conference; December 7-10; Miami, Florida p. 9. [doi: 10.1109/wsc.2008.4736059]
63. R: the R project for statistical computing. R Foundation for Statistical Computing. URL: http://www.R-project.org/ [accessed 2019-03-13]
64. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. Adm Policy Ment Health 2011 Jan 22;38(1):44-53 [FREE Full text] [doi: 10.1007/s10488-010-0314-z] [Medline: 20967495]
65. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
66. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies. Qual Health Res 2016 Jul 10;26(13):1753-1760. [doi: 10.1177/1049732315617444]
67. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
68. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research. Int J Acad Med 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
69. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May 12;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
70. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS One 2020 Jan 15;15(1):e0225695 [FREE Full text] [doi: 10.1371/journal.pone.0225695] [Medline: 31940347]

XSL•FO
RenderX

Edited by R Kukafka, G Eysenbach; submitted 12.02.21; peer-reviewed by PCI Pang, A Teles; comments to author 10.03.21; revised version received 26.03.21; accepted 02.04.21; published 06.05.21.

Please cite as:
Areán PA, Pratap A, Hsin H, Huppert TK, Hendricks KE, Heagerty PJ, Cohen T, Bagge C, Comtois KA
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;23(5):e27918
URL: https://www.jmir.org/2021/5/e27918
doi:10.2196/27918
PMID:33955838

©Patricia A Areán, Abhishek Pratap, Honor Hsin, Tierney K Huppert, Karin E Hendricks, Patrick J Heagerty, Trevor Cohen, Courtney Bagge, Katherine Anne Comtois. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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

Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions

Akira-Sebastian Poncette[1,2], MD; Maximilian Markus Wunderlich[2], MSc; Claudia Spies[1], MD, PhD; Patrick Heeren[1,2], MSc; Gerald Vorderwülbecke[1], MD; Eduardo Salgado[1,2], MD; Marc Kastrup[1], MD, PhD; Markus A Feufel[3], MSc, PhD; Felix Balzer[1,2], MD, MSc, PhD

[1]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[3]Department of Psychology and Ergonomics (IPA), Division of Ergonomics, Technische Universität Berlin, Berlin, Germany

Corresponding Author:
Felix Balzer, MD, MSc, PhD
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin
Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450 ext 651166
Email: felix.balzer@charite.de

Abstract

Background: As one of the most essential technical components of the intensive care unit (ICU), continuous monitoring of patients' vital parameters has significantly improved patient safety by alerting staff through an alarm when a parameter deviates from the normal range. However, the vast number of alarms regularly overwhelms staff and may induce alarm fatigue, a condition recently exacerbated by COVID-19 and potentially endangering patients.

Objective: This study focused on providing a complete and repeatable analysis of the alarm data of an ICU's patient monitoring system. We aimed to develop do-it-yourself (DIY) instructions for technically versed ICU staff to analyze their monitoring data themselves, which is an essential element for developing efficient and effective alarm optimization strategies.

Methods: This observational study was conducted using alarm log data extracted from the patient monitoring system of a 21-bed surgical ICU in 2019. DIY instructions were iteratively developed in informal interdisciplinary team meetings. The data analysis was grounded in a framework consisting of 5 dimensions, each with specific metrics: alarm load (eg, alarms per bed per day, alarm flood conditions, alarm per device and per criticality), avoidable alarms, (eg, the number of technical alarms), responsiveness and alarm handling (eg alarm duration), sensing (eg, usage of the alarm pause function), and exposure (eg, alarms per room type). Results were visualized using the R package ggplot2 to provide detailed insights into the ICU's alarm situation.

Results: We developed 6 DIY instructions that should be followed iteratively step by step. Alarm load metrics should be (re)defined before alarm log data are collected and analyzed. Intuitive visualizations of the alarm metrics should be created next and presented to staff in order to help identify patterns in the alarm data for designing and implementing effective alarm management interventions. We provide the script we used for the data preparation and an R-Markdown file to create comprehensive alarm reports. The alarm load in the respective ICU was quantified by 152.5 (SD 42.2) alarms per bed per day on average and alarm flood conditions with, on average, 69.55 (SD 31.12) per day that both occurred mostly in the morning shifts. Most alarms were issued by the ventilator, invasive blood pressure device, and electrocardiogram (ie, high and low blood pressure, high respiratory rate, low heart rate). The exposure to alarms per bed per day was higher in single rooms (26%, mean 172.9/137.2 alarms per day per bed).

Conclusions: Analyzing ICU alarm log data provides valuable insights into the current alarm situation. Our results call for alarm management interventions that effectively reduce the number of alarms in order to ensure patient safety and ICU staff's

XSL•FO
RenderX

work satisfaction. We hope our DIY instructions encourage others to follow suit in analyzing and publishing their ICU alarm data.

(J Med Internet Res 2021;23(5):e26494) doi:10.2196/26494

KEYWORDS

digital health; patient monitoring; intensive care unit; technological innovation; data science; alarm fatigue; alarm management; patient safety; ICU; alarm system; alarm system quality; medical devices; clinical alarms

Introduction

Background

In intensive care units (ICUs), monitoring of patients' physiologic parameters has significantly improved patient safety by alerting ICU staff through a visual or audible alarm [1] when a parameter deviates from the preset range (eg, apnea, sensor detachment). However, as one of the most digitized health care areas with a rising number of novel devices with their own alarms, the sheer number of alarms regularly overwhelms ICU staff. Some studies document more than 700 alarms per patient per day on average [2].

Being exposed to so many alarms can leave ICU staff alarm fatigued, a condition characterized by a desensitization to alarms, which can make ICU staff react inadequately (eg, responding with delay, turning down the alarm volume, turning alarms off) [3,4]. Due to the COVID-19 pandemic, this condition has been further exacerbated (eg, through the utilization of anesthesia ventilators in the ICU) [5]. Excessive alarms not only induce stress and distraction in ICU staff [6,7] but also directly impair patient recovery [8]. Ultimately, it can threaten patients' lives when ICU staff misses alarms or responds with delay. This is shown by the Joint Commission's sentinel event database, which lists 98 incidents between 2009 and 2012 that were related to alarms, of which 80 resulted in a patient's death [9]. Reporting to the database is voluntary, which likely makes this a conservative estimate. For 2020, the ECRI Institute listed an alarm-related hazard among their Top 10 Health Technology Hazards [10]. While in the United States, the Joint Commission declared it a national patient safety goal to reduce the harm associated with clinical alarm systems from 2014 onwards [11]; there is no such official endeavor in Germany.

One way to reduce harm associated with clinical alarms is alarm management, which aims to reduce the number of unnecessary alarms (that is, false, nonactionable, and avoidable technical alarms [12]) with the assumption that this reduces the overall number of alarms and thereby alleviates the staff's alarm fatigue. Traditional alarm management approaches that have been proven to reduce the overall number of alarms include the recommendation to mute alarms while examining a patient [13], introduce a delay between measuring and alarming [14], use individual thresholds for each patient instead of the monitoring device's default [15], turn off arrhythmia alarms that are not life threatening, and change electrocardiogram (ECG) leads on a daily basis [16].

To be most effective, alarm management should be adjusted to the specific conditions of each ICU [12,17]. A thorough analysis of the sociotechnical system of the ICU is necessary to sufficiently customize respective interventions. These efforts include the analysis of the alarm log data (eg, when an alarm occurred and by which device) [12,18]. Currently there is no software solution commercially available that addresses analysis of patient monitoring data.

Aim

Our aim is to develop do-it-yourself (DIY) instructions targeted at technically versed ICU staff (physicians and nurses) for self-analysis of patient monitoring alarm data, including an illustrative, complete, and repeatable analysis of device alarm data of an ICU's patient monitoring system. The application of the DIY instructions should help their users to identify patterns and trends in the alarm data and enable them to generate ideas on how the overall alarm frequency (and subsequently alarm fatigue) might be reduced.

Methods

Ethics Approval

The ethical approval for this study was granted by the Ethics Commission of the Charité – Universitätsmedizin Berlin (EA1/127/18).

Setting and Design

We conducted the study in a surgical ICU of a German university hospital. The unit consists of 21 beds in 15 rooms in which mainly patients after abdominal or neurosurgical operations are treated. The patient monitoring and alarm system used at the time of the study was the Philips IntelliVue patient monitoring system (MX800 software version M.00.03; MMS X2 software version H.15.41-M.00.04; Philips, Amsterdam, Netherlands) with bedside monitors, 3 client monitors summarizing 2-3 rooms, a central station (software version B), and 2 large hallway monitors displaying all 21 patients. Standard monitoring included oxygen saturation (SpO_2), heart rate, invasive (IBP) or noninvasive blood pressure (NIBP), and temperature. Within the Clinical Alarm Capability Maturity Model by Welch et al [19], the ICU was in the first stage at the time of this study, described as having many nonactionable alarms for unknown reasons, approaching alarm management ad hoc and not having or not consulting data to support change [19]. Accordingly, there was no hospital-wide consensus on alarm management.

We used an observational study design, which included retrospective data analysis of the patient monitoring alarm data. The DIY instructions were iteratively developed in informal interdisciplinary team meetings within the research group between February 2019 and November 2020 and adapted by the lessons learned from our own data analysis.

XSL•FO
RenderX

Data Collection and Deidentification

We manually collected clinical audit logs (which include the alarm data) 3 times during 2019 (in winter, summer, and autumn) via a USB stick from the central patient monitoring device in the ICU as previously described by others. The clinical audit log consists of the time, bed number, alarm type (ie, parameter, device, alarm criticality), and alarm handling (eg, threshold adjustments, use of the pause function). Each log file contains data from 31 days.

No actual patient-identifying data elements were collected. For further deidentification, dates were shifted into the future by a pseudorandom offset for all patients; the bed number was replaced by a pseudonym. Day and night rhythm, weekends, the season, and the bed characteristic (double room, single room) were not affected by this process. The deidentified raw data can be retrieved from an open data repository [20].

Data Analysis Framework

Overview

We organized our data analysis in a framework based on suggestions by Hüske-Kraus et al [12], who introduced quality dimensions along with metrics of an alarm system. Each dimension summarizes multiple metrics (Table 1).

Table 1. Data analysis framework applied in this study in line with the quality dimensions introduced by Hüske-Kraus et al [12] and including metrics suggested by Hüske-Kraus et al [12] as well as metrics suggested by other sources for each dimension, wherever possible.

Quality dimension	Definition	Metrics used in this study
Alarm load	Metrics related to the number of alarms	Alarms per bed per day, frequency of individual alarms, alarms per device, alarms per criticality (red, yellow, and blue; ie, alarm at high criticality, alarm at medium criticality, and technical alarm at low criticality, respectively), average temporal distribution of alarms and alarm flood conditions (10 or more alarms occurring within 10 minutes) [18]
Avoidable alarms	False-positive alarms, nonactionable alarms, and technical alarms	Technical alarms per bed per day, technical alarms per device
Responsiveness and alarm handling	Alarm duration, response time, muting of alarms, and corrective actions	Duration of alarms
Sensing	The quality of the technical infrastructure, such as consumable, overmonitoring, and undermonitoring	Average usage of the alarm pause function per bed per day, proper pause-to-pause ratio [12], redundant monitoring of physiological parameters
Exposure	How alarms are distributed in the unit	Average alarm frequencies per room and per bed per room type, number of beds issuing more alarms than the average

Data Analysis

We cleaned and analyzed the data with R [21] in combination with the packages *dplyr* [22], *tidyr* [23], and *stringr* [24]. We used the package *lubridate* [25] for date and time information and the package *ggplot2* [11] for the visualizations. The log entries were structured in 4 columns: Time, Bedname, Action, Devicename. New variables (eg, the time an alarm was generated and its criticality) were extracted for each log entry from the information contained in the column Action. In total, the alarm logs contain data from 93 days.

Alarm Load

Visualizing the frequency of individual alarm parameters helps to identify "bad actors" — alarms that occur much more frequently than others [8] while investigating the number of alarms each medical device issues — can help to prioritize alarm management interventions. The metric "alarms per bed per day" alone is not necessarily an indicator of the alarm load on respective ICUs and should be accompanied with information such as the criticality of the alarms (red, yellow, blue), the frequency of individual alarms, the frequency of alarms per device, the number of alarm flood conditions, and a temporal perspective. To conduct device-related analyses, we assigned each alarm parameter to 1 out of 7 devices (ventilator, ECG, IBP, intracranial pressure, temperature, NIBP, and pulse oximetry [SpO_2]). Technical alarms included alarms with a blue criticality and general monitoring device–related alarms (such as missing patient information, low batteries, or interrupted arterial blood pressure measurements). Devices, where the absolute cumulative frequency of alarms was less than 500 in the dataset were only included in the overall count of alarms (Multimedia Appendix 1).

To visualize the average distribution of alarms across 24 hours, we calculated the average number of alarms of each 1-minute bin between 12:00 am and 11:59 pm for the 3 devices issuing the most alarms and applied the scatterplot smoothing function of the R package ggplot2 [26].

Alarm flood conditions are described as situations in the ICU where multiple alarms are triggered within a short time frame. Metrics related to alarm flood conditions provide information that allow an additional perspective on the alarm load of an ICU and take an acute overload of ICU staff by alarms into account [18]. We grouped the data per bed and split each bed's data into 10-minute bins starting from the date-time stamp of the first log entry. All bins containing 10 or more alarms were counted as an alarm flood, as previously described [18].

XSL•FO
RenderX

Avoidable Alarms

Avoidable alarms are defined as nonactionable (including false positive alarms) or technical [12]. Since the alarm log data lack information on whether an alarm was true or false or if it was followed by a therapeutic intervention, we cannot provide metrics such as the positive predictive value of alarms. However, most technical alarms, whether they were responded to with an intervention or not, are avoidable nonetheless [12]. Therefore, we report the average number of technical alarms per bed per day as well as the individual frequency of technical alarms.

Responsiveness and Alarm Handling

We visualize the median alarm duration per medical device in absolute seconds and over the course of 24 hours for the 3 devices issuing the most alarms. The duration of an alarm was defined as the time difference between the timestamp of the log entry of a generated alarm and that of a terminated alarm. We opted for this method, because only the "true" time of generated alarms is documented but not the "true" time of terminated alarms, where only the time of the log entry is provided.

Sensing

Using the pause function of the monitoring devices is argued to prevent unnecessary alarms [16,27]. In the investigated ICU, the alarm pause function suspends all alarms for up to 3 minutes or until it is actively terminated by the medical team. This helps prevent alarms that would be triggered when, for example, a patient is being shifted from one position to another during a physical examination. A responsible use of the pause function demands its active termination to avoid undermonitoring of the patient due to the suspended alarms for the remainder of the pause function [12] once the health care provider leaves the patient. Hence, an alarm pause can be considered a proper pause if its duration is shorter than the monitor's default pause duration. The metric "proper pause-to-pause ratio" indicates the ratio of proper pauses to all pauses [12]. We count pauses that were re-enabled within 3 minutes after their termination as one continuous pause, since the default length might not have been sufficient for the bedside procedure.

We consider an overmonitoring of patients to be indicated when parameters that monitor the same physiological event, but stem from 2 different medical devices, are routinely issuing alarms.

Exposure

The present ICU has 2 different room types: a single bedroom and a double bedroom. We aimed to find out whether a bed in a single bedroom produces more, less, or equally as many alarms as a bed in a double bedroom and whether this differs depending on the alarm criticality. In total, there are 12 beds in double bedrooms and 9 beds in single bedrooms. The average alarm frequencies per bed per room type were calculated by dividing the number of alarms per room type by 12 or by 9, respectively.

Results

Overview

The results section is clustered into 2 parts: First, we provide the DIY instructions for self-analysis of alarm data from the patient monitoring system; second, we present the results of the illustrative alarm data analysis conducted on data from our ICU.

DIY Instructions

Step-by-Step Procedure

The DIY instructions for self-analysis of the ICU's patient monitoring alarm data consist of 6 steps: (1) define or re-evaluate metrics, (2) collect data, (3) analyze data, (4) visualize metrics, (5) present to staff and set goals together, and (6) design and implement interventions (see Figure 1).

Figure 1. Feedback loop regarding do-it-yourself (DIY) instructions for self-analysis of patient monitoring alarm data in the intensive care unit.



XSL•FO
RenderX

Define or Re-Evaluate Metrics

In an interdisciplinary team consisting of ICU physicians and nurses, alarm data metrics should be defined. We recommend to initially include all metrics presented in the aforementioned data analysis framework in order to get an accurate and complete picture of the alarm situation of the respective unit. After the first iteration of the feedback loop, new metrics may be added or existing ones modified.

Collect Data

Regular collection of patient monitoring data is crucial to conduct reliable data analyses, especially if interventions are being conducted at the respective ICU. The monitoring central station at our ICU stores data for up to 90 days; hence, every 90 days, data have to be manually extracted from the system [28].

Analyze Data

We provide the fully annotated R scripts that we used to conduct the alarm data analysis to enable even beginners in R to do likewise. Further explanations can be found in the Results section and in the scripts [20].

Visualize Metrics, Present to Staff, and Set Goals

Visualizations should be summarized in a clear and intuitive format (eg, using a presentation program) and discussed with ICU staff. Together, realistic goals should be set for each parameter and possible interventions deduced. On a quarterly basis, this feedback loop should be started from the beginning. The R-Markdown file on GitHub can also be used to create comprehensive alarm reports including all metrics and visualizations reported in this paper [20].

Design and Implement Interventions

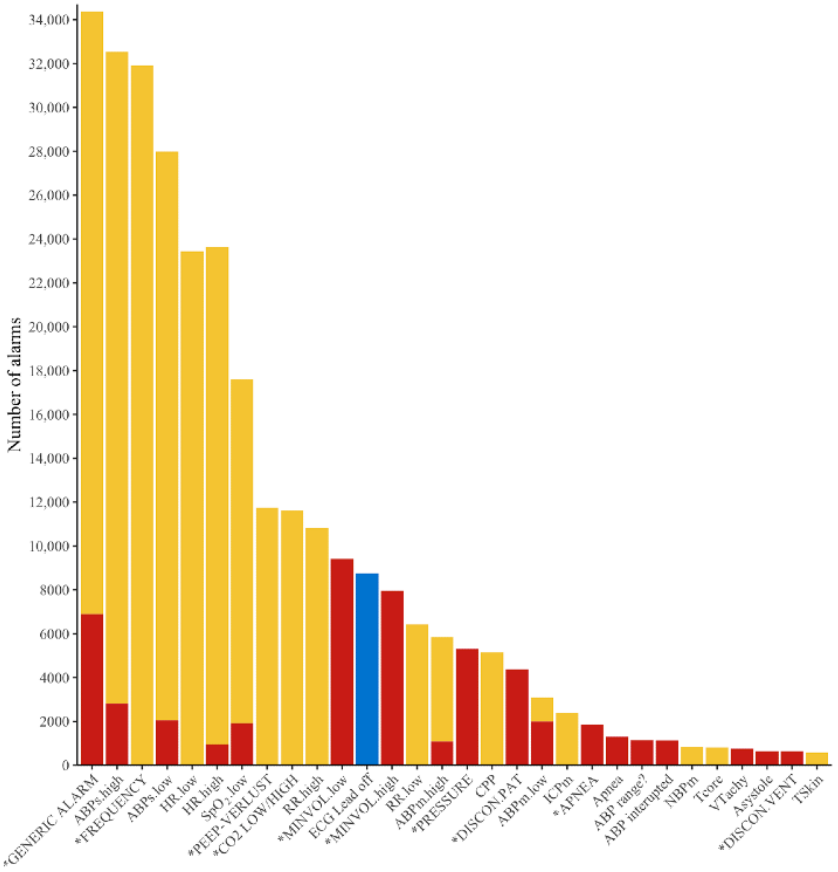
Potential interventions are deduced from the data analysis, the visualizations, and the interdisciplinary goal setting. ICU staff as the end users of the patient monitoring system should be actively involved in the intervention design and rotate regularly with new medical staff from the ICU to assess clinical relevance more reliably. Possible interventions could include adjusting the default alarm thresholds for the top 10 alarm parameters, focusing on customizing these parameters more frequently, introducing new ECG electrodes or skin preparation routines, or providing staff training (eg, on how to use the alarm pause function). Further interventions are elaborated elsewhere [29,30].

Data Analysis

Alarm Load

The analyzed alarm log data set contained, on average, 152.5 (SD 42.2) alarms per bed per day. Most alarms were type yellow (mean 120.3/152.5, SD 37.15 per day; 79%), followed by type red (mean 27.5/152.5, SD 9.37 per day; 18%). Few alarms were type blue (mean 4.6/152.5, SD 2.75 per day; 3%). The 5 most frequent alarms were a generic alarm of the ventilator, invasive systolic blood pressure (high and low), high respiratory rate issued by the ventilator, and low heart rate (see Figure 2).

Figure 2. Frequency of individual alarm parameters within 93 days. The colors correspond to the alarm criticalities (red, yellow, and blue). *: ventilator arm; ABPs: systolic arterial blood pressure; ECG: electrocardiogram; FREQUENCY: ventilator alarm indicating that the upper respiratory rate threshold has been exceeded; HR: heart rate; RR: respiratory rate derived from the ECG (see Multimedia Appendix 2 for all abbreviations).



XSL•FO
RenderX

After each alarm parameter was assigned to the corresponding medical device, it was evident that the ventilator generates the most alarms, followed by IBP and ECG (see Figure 3).

When put into a temporal perspective, the average distribution of alarms across 24 hours for the 3 devices that issue the most alarms shows a downward trend, with most alarms being issued in the morning shift and fewest during the night (see Figure 4). The mean number of alarms per minute per medical device during the morning, afternoon, and night shifts were: 1.09 (SD 0.2), 1.0 (SD 0.18), and 0.71 (SD 0.17) for the ventilator; 0.65 (SD 0.13), 0.56 (SD 0.1), and 0.45 (SD 0.09) for IBP; and 0.57 (SD 0.12), 0.51 (SD 0.1), and 0.45 (SD 0.09) for ECG, respectively.

Figure 3. Alarms from medical devices within 93 days subdivided into the criticality levels (red, yellow). ECG: electrocardiogram; IBP: invasive blood pressure; ICP: intracranial pressure; NIBP: noninvasive blood pressure; SpO_2: oxygen saturation.

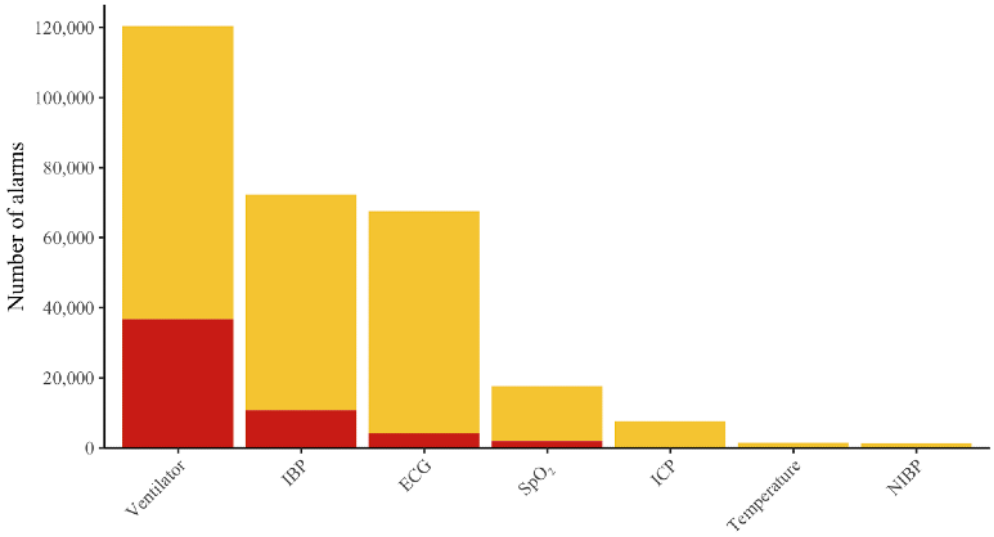
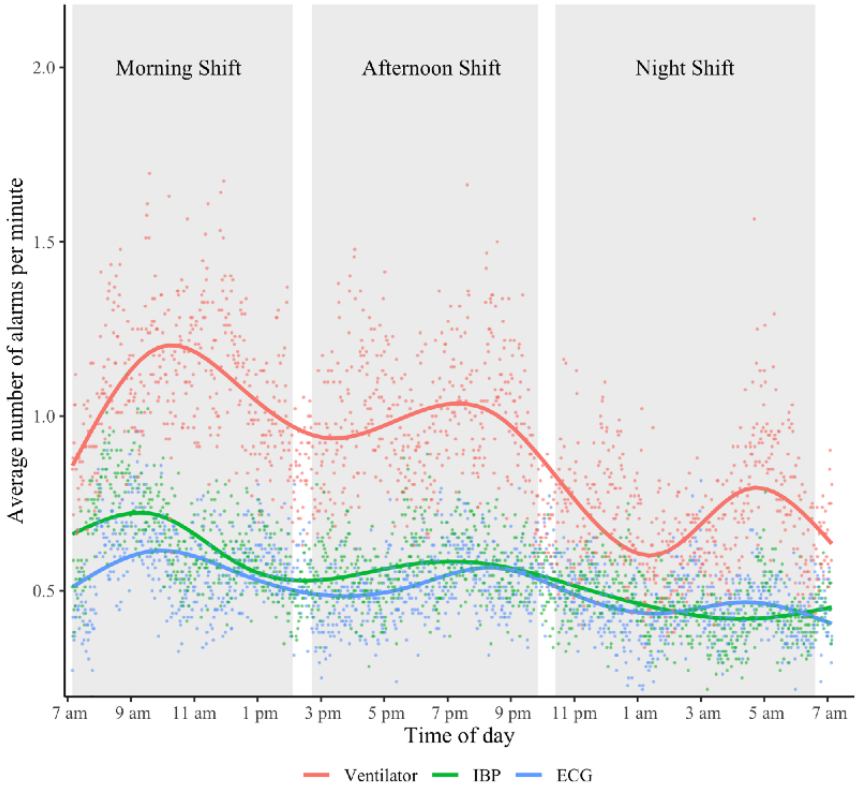


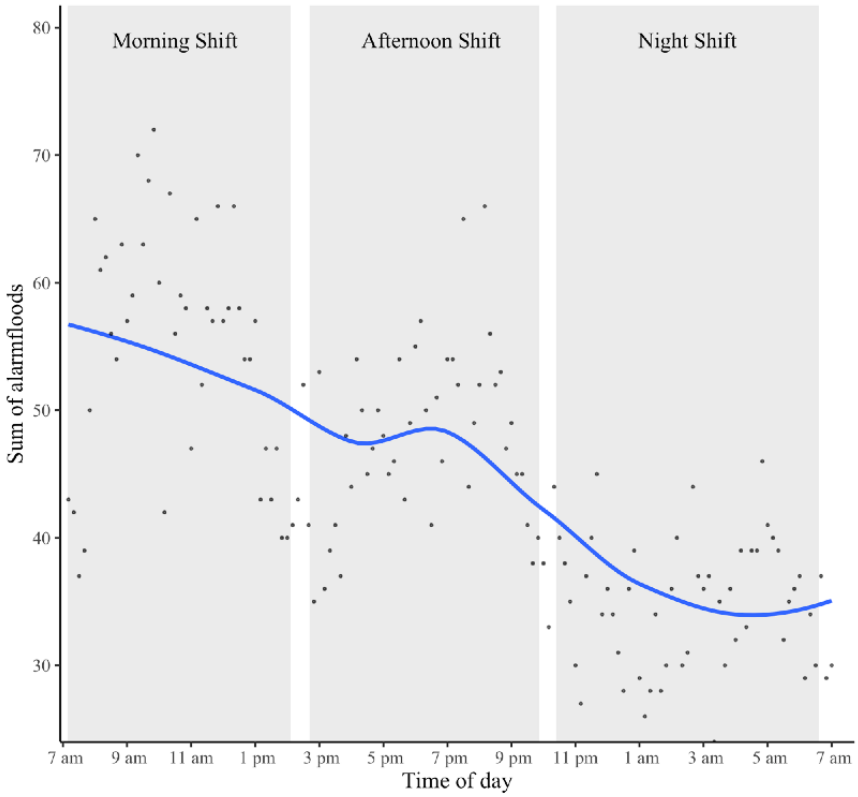
Figure 4. Average distribution of alarms across 24 hours. The white spaces between the grey bars (ie, shifts) visualize handover periods. Each dot shows the average alarm frequency of 1 minute for the specified device. The line for each device is calculated by ggplot2's smoothing function and represents a generalized additive model of the distribution (with the formula y ~ s(x, bs = "cs"). It serves to aid in detecting trends in the data. ECG: electrocardiogram; IBP: invasive blood pressure.



XSL•FO
RenderX

In total, 6468 alarm flood conditions occurred (mean 69.55, SD 31.0 per day; median 63; range 22-194), of which 5289 (82%) were comprised of between 10 and 20 alarms, 1012 (16%) between 20 and 40 alarms, and 159 (2%) between 40 and 100 alarms within 10-minute intervals. The temporal visualization over 24 hours shows a general downward trend with peaks in the morning and afternoon shifts (see Figure 5).

Figure 5. Temporal distribution of alarm flood conditions over 24 hours. Each dot indicates the sum of all alarm flood conditions that were initiated at the respective time of day in 10-minute intervals. For example, the first dot on the far left indicates that 43 alarm floods occurred between 7:10 and 7:20 AM across all days in the data. The blue line is a local regression, calculated by ggplot2's smoothing function (formula: y ~ x). The white spaces between the grey bars (ie, shifts) visualize handover periods.



Avoidable Alarms

In total, 10,846 technical alarms (red, yellow, and blue) are documented. This equals 5.6 (SD 2.8) technical alarms per bed per day, on average. With 8746 alarms (all blue), the ECG produced the most technical alarms (ECG lead fallen off), followed by IBP (1342 red alarms) and alarms related to the module cable connection (167 blue alarms).

Responsiveness and Alarm Handling

The alarm duration showed a distribution that is strongly skewed to the right (for clinical alarms: mean 109.3, SD 6109.15 seconds; median 8 seconds; range 0-2,291,314 seconds; for technical alarms: mean 221.5, SD 4,898 seconds; median 7 seconds; range 0-403,440 seconds), which is why we used the median as the measure of the center for further analyses and plots. Additionally, because some durations were calculated to be unrealistically high (in some instances, multiple days), we treat all durations longer than 8 hours (approximately the length of one shift) as outliers and do not include them in the analyses.

Median alarm durations of yellow and red alarms were similar (8 seconds; range 0-22,048 seconds; range 0-28,531 seconds), while the median alarm duration of blue alarms was slightly longer (9 seconds; range 0-26,049 seconds). Regarding devices, the median duration of alarms by NIBP was the longest (64 seconds; range 0-3845 seconds), followed by temperature (26 seconds; range 0-7796 seconds), SpO_2 (16 seconds; range 0-27,580 seonds), and IBP (14 seconds; range 0-28,531 seconds). ECG and the ventilator recorded the lowest median alarm duration (4 seconds; range 0-27,579 seconds and 7 seconds; range 0-22,048 seconds, respectively). Visualizing the alarm durations over 24 hours shows that the difference across devices was relatively stable over the course of an average day (see Figure 6). However, Figure 7 shows that there were substantial differences between the median duration to red and yellow ECG alarms.

XSL•FO
RenderX

Figure 6. Median alarm duration of the 3 medical devices that issue most alarms over 24 hours. Each dot represents the median alarm duration for each minute of the day of the respective device. The line for each device is based on ggplot2's smoothing function and represents a generalized additive model of the distribution (with the formula y ~ s(x, bs = "cs"). The white spaces between the grey bars (ie, shifts) visualize handover periods. ECG: electrocardiogram; IBP: invasive blood pressure.

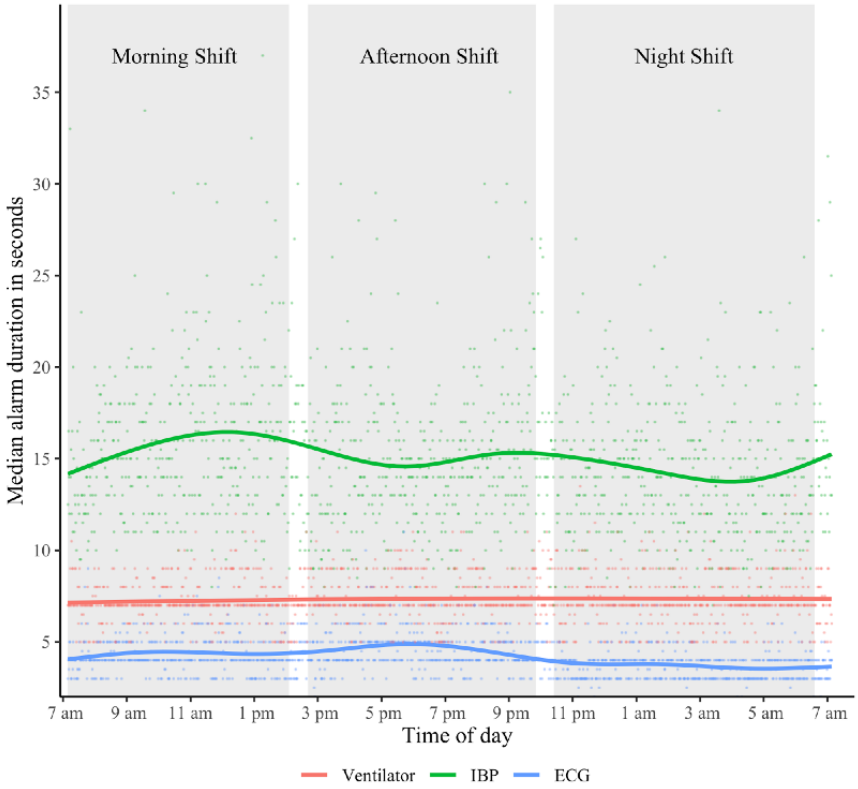
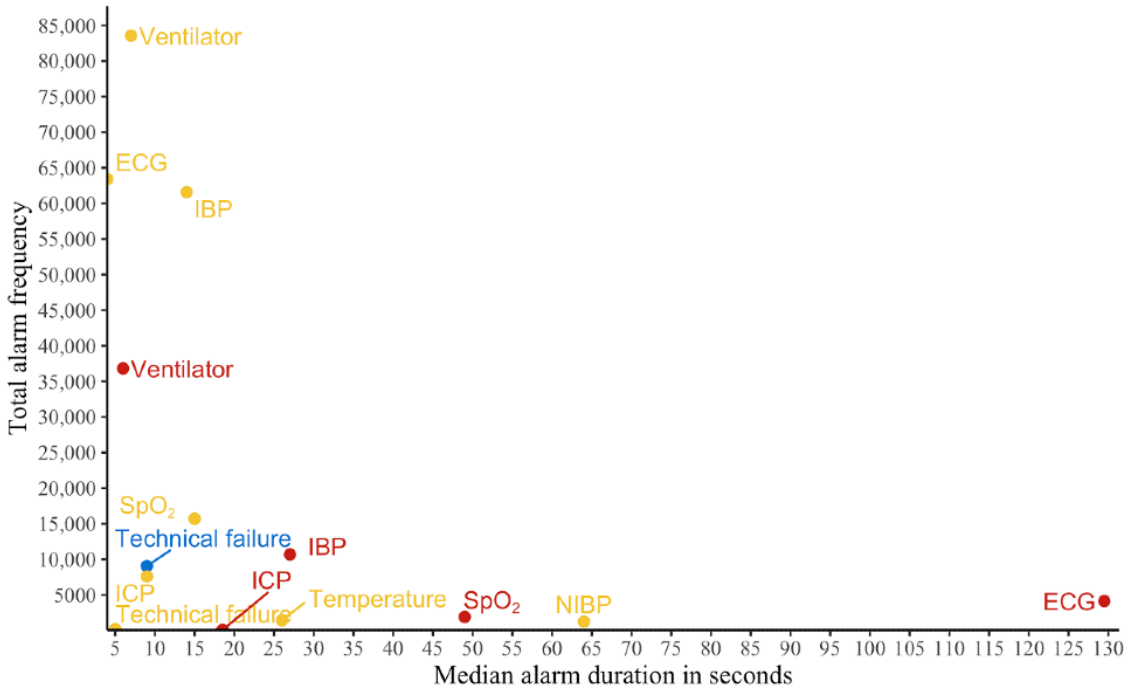


Figure 7. The median alarm duration from 8 medical devices plotted against the total number of alarms issued by the respective device. The colors correspond to the alarm criticalities (red, yellow, and blue). ECG: electrocardiogram; IBP: invasive blood pressure; ICP: intracranial pressure; NIBP: noninvasive blood pressure; SpO_2: oxygen saturation.



XSL•FO
RenderX

Sensing

On average, the alarm pause function was applied 10.86 (SD 2.6) times per bed per day. Of all pauses that were started, 92% (14,719/16,002) were not actively terminated but lasted for their default maximum length of 3 minutes and therefore do not qualify as proper pauses [10]. The ICU's proper pause-to-pause ratio is 0.09:1.

Of the ECG alarms, 16% (10,821/67,518) indicated a high respiratory rate, which amounted to 4% (10,821/297,830) of all alarms recorded in the data set, while the alarms from the ventilator related to a high respiratory rate covered another 11% (31,911/297,830) of all alarms. Additionally, both device groups had similar numbers of life-critical apnea alarms. This suggests an overmonitoring of the respiratory rate.

Exposure

A bed located in a single bedroom had, on average, 26% (172.9/137.2) more alarms per day than a bed located in a double bedroom. There were 32% more red alarms per bed in a single bedroom than in a double bedroom (2972.9/2250.3) and 25% more yellow alarms per bed in a single bedroom than in a double bedroom (12,638.8/10,105.2).

The calculated average alarms per bed per day yields 152.5 alarms (SD 42.2). On average, 36% (7.6/21, SD 1.6) of the 21 beds exceeded the units average every day, issuing on average 69% of all daily alarms (2199.9/3202.4, SD 651.2).

Discussion

Principal Findings

We aimed to provide technically versed ICU staff with a framework and the tools to conduct a self-analysis of patient monitoring alarm data in order to help them assess their unit's alarm situation, inspect potential root causes of excessive alarms, and derive alarm management interventions that might help to remedy alarm fatigue. Our framework consists of 6 steps that should be iteratively applied: (1) define or re-evaluate metrics, (2) collect data, (3) analyze data, (4) visualize metrics, (5) present to staff and set goals together, and (6) design and implement interventions. We designed the framework to be useful independent of the ICU's specialization (eg, COVID-19 units, neonatal ICUs, pediatric ICUs). In our observational study, we illustrated how the alarm log data of a large German ICU can be analyzed and how alarm metrics can be visualized using the scripts that we provide (steps 2 and 3 of the framework, respectively [20]). The data analysis was structured according to the aforementioned quality dimensions [12]: The alarm load was quantified by 152.5 (SD 42.2) alarms per bed per day on average, issued mostly by the ventilator, IBP measurement, and ECG in the morning shifts (ie, high/low blood pressure, high respiratory rate, high/low heart rate). Alarm flood conditions also mostly occurred in the morning shifts with, on average, 69.55 (SD 31.12) per day. With regard to avoidable alarms, technical alarms were mostly issued by the ECG (ie, lead fallen off). The dimension "responsiveness and alarm handling" included the metric "alarm duration." The calculation yielded a median duration of 8 (range 0-2,291,314) seconds for clinical alarms and 7 (range 0-403,440) seconds for technical alarms. Regarding "sensing," the alarm pause function is, on average, applied 10.86 (SD 2.6) times per bed per day, and in 92% (14,719/16,002) was not actively terminated, resulting in a proper pause-to-pause ratio of 0.09:1. The "exposure" to alarms per bed per day was higher in single rooms (26%, mean 172.9/137.2 alarms per day per bed). Most alarms were, on average, issued by 7.6 of 21 beds (36%).

Alarm Metrics in Perspective

Cvach et al [29] suggested that most ICU patients have less than the average number of alarms per bed per day while a few have more than that. In their data analysis of adult telemetry, 19% (n=3) exceeded the unit's average on a single day; in our data analysis, on average, 36% (n=7.6) exceeded the unit's average.

Our data analysis shows that beds located in a single bedroom have a higher alarm load compared to a bed in a double bedroom. Further analysis of alarm data at the patient level could reveal whether the alarm load depends on the severity of the patient's illness. Upon presenting our results to ICU staff, they noted that patients in delirium are often treated in single bedrooms. Since delirium is a condition that can lead to erratic movements [31], this might explain the larger number of alarms coming from single bedrooms (eg, due to disconnected ECG leads). This anecdote highlights the importance of presenting the results of the data analysis to ICU staff, as suggested in Figure 1.

Slow response times can be an indicator of alarm fatigue [32]. However, response times to alarms can be slow for other reasons than alarm fatigue alone, such as the unit's floor layout and policies [17] or the staff members' individual personality traits [33]. Similarly, response times can be fast for other reasons than a well-functioning organization: ICU staff might be so severely desensitized to alarms that they start terminating them blindly, without properly evaluating the patient's situation [12]. Although response time and alarm duration are related, they are not the same. Response time describes the time between the generation of an alarm and its manual termination. Alarm duration describes the time between the generation of an alarm and any termination of that alarm (such as the auto-termination of alarms if an alarm with a higher priority is issued). Our data did not include information on whether an alarm was manually terminated or not, which is why we focused on analyzing the alarm duration.

Although the pause function of the monitoring devices can prevent unnecessary alarms [14,26], it should be used responsibly by terminating it before leaving the patient. Our data yielded a proper pause-to-pause ratio of 0.09:1, meaning that most pauses that were started were not actively terminated but lasted for their default maximum length of 3 minutes. This ratio is far from Hu¨ske-Kraus et al's [12] ideal ratio of 1, where all pauses would be a "proper pause." If it is indeed the case that health care providers leave patients with the alarm pause still engaged, then this aspect should be included in future staff training to promote a responsible use of the alarm pause function. However, other reasons for this pattern in the data should be considered as well. For example, sometimes even the maximum duration of an alarm pause is not long enough (eg, when ICU staff are in the middle of an intervention on the other

XSL•FO
RenderX

side of the bed, unable to reach the patient monitor, watching the pause automatically disengage). Hence, this nonideal proper pause ratio does not necessarily represent carelessness of ICU staff but could hint towards the maximum default length of alarm pauses as being too short.

Limitations

Quantification of the alarms does not reflect whether alarm fatigue is an issue in the respective ICU. In order to evaluate alarm fatigue as a complex sociotechnical phenomenon, the data analysis should be accompanied by a qualitative study (eg, by staff interviews or alarm fatigue surveys) [34]. Our applied grouping of metrics into dimensions is based on available literature, not delimited, and to some extent arbitrary. For example, the “alarm pause” metric could be assigned to the dimension “alarm handling” or the “alarm flood” metric to the dimension “exposure.” The software version of our monitoring system does not log technical alarms with a low priority (soft inoperable alarms), and we did not include alarms from every medical device that issues alarms (eg, perfusion pump alarms are not included). Hence, the metrics reported underestimate the actual alarm load of the unit.

Conclusion

We demonstrated that basic data analysis skills can help generate valuable insights for designing alarm management interventions and how alarm data analyses might be embedded in an overarching framework that guides in developing such interventions. We hope the presented DIY instructions and the alarm processing and visualization scripts accompanying this publication will be helpful to other intensivists and researchers and spur the publication of many ICUs’ alarm data and lessons learned from their alarm management efforts.

Acknowledgments

We express our gratitude to Julia Liehr, medical technician of the Charité CFM Facility Management GmbH, for supporting us with her technical expertise regarding patient monitoring, as well as Dirk Hüske-Kraus at Philips Healthcare for his valuable feedback on our manuscript. We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Funds of Charité – Universitätsmedizin Berlin. AP is a participant in the Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health.

Authors' Contributions

MK identified the need to improve the alarm situation in the ICU and initiated the study together with AP. The study was conceived by AP, CS, FB, MF, and MW. AP conducted data acquisition, supported by ES and GV. MW and AP analyzed the data, supported by PH as a computer scientist. AP wrote the manuscript, supported by MW. FB supervised all parts of the study. All authors critically reviewed and approved the manuscript. The article was extracted from the M.S. thesis of MW.

Conflicts of Interest

CS received public funding from the Stifterverband and the Einsteinstiftung, and received funding from different companies for the Leopolidna Meeting 2020 payed to the Charité - Universitätsmedizin Berlin. All other authors declare no conflicts.

Multimedia Appendix 1
Patient monitoring device group assignments.
[XLSX File (Microsoft Excel File), 11 KB - jmir_v23i5e26494_app1.xlsx]

Multimedia Appendix 2
Alarm notification abbreviations.
[XLSX File (Microsoft Excel File), 31 KB - jmir_v23i5e26494_app2.xlsx]

References

1. Moreno RP, Rhodes A, Donchin Y, European Society of Intensive Care. Patient safety in intensive care medicine: the Declaration of Vienna. Intensive Care Med 2009 Oct;35(10):1667-1672 [FREE Full text] [doi: 10.1007/s00134-009-1621-2] [Medline: 19697007]
2. Jones K. Alarm fatigue a top patient safety hazard. CMAJ 2014 Feb 18;186(3):178 [FREE Full text] [doi: 10.1503/cmaj.109-4696] [Medline: 24418978]
3. Sendelbach S, Funk M. Alarm Fatigue. AACN Advanced Critical Care 2013;24(4):378-386. [doi: 10.1097/nci.0b013e3182a903f9]
4. Wilken M, Hüske-Kraus D, Klausen A, Koch C, Schlauch W, Röhrig R. Alarm Fatigue: Causes and Effects. Stud Health Technol Inform 2017;243:107-111. [Medline: 28883181]
5. Xi AS, Chang MG, Bittner EA. Rapid establishment of an ICU using anesthesia ventilators during COVID-19 pandemic: lessons learned. Crit Care 2020 Jun 30;24(1):388 [FREE Full text] [doi: 10.1186/s13054-020-03107-x] [Medline: 32605580]

XSL•FO
RenderX

6. Poncette A, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical Requirements of Future Patient Monitoring in the Intensive Care Unit: Qualitative Study. JMIR Med Inform 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: 10.2196/13064] [Medline: 31038467]
7. Bach TA, Berglund L, Turk E. Managing alarm systems for quality and safety in the hospital setting. BMJ Open Qual 2018 Jul 25;7(3):e000202 [FREE Full text] [doi: 10.1136/bmjoq-2017-000202] [Medline: 30094341]
8. Clinical Alarm Management Compendium. AAMI Foundation. 2015. URL: http://essentialhospitals.org/wp-content/uploads/2015/11/Alarm_Compendium_2015.pdf [accessed 2021-05-08]
9. Medical device alarm safety in hospitals. Joint Commission Sentinal Event Alert. 2013 Apr 08. URL: https://www.jointcommission.org/-/media/deprecated-unorganized/imported-assets/jcr/assets/sea_50_alarms_4_5_13_final1pdf.pdf [accessed 2021-05-18]
10. 2020 Top 10 Health Technology Hazards Executive Brief. ECRI. URL: https://www.ecri.org/landing-2020-top-ten-health-technology-hazards [accessed 2021-05-08]
11. National Patient Safety Goals Effective January 2018. The Joint Commission. 2018. URL: https://www.jointcommission.org/-/media/tjc/documents/standards/national-patient-safety-goals/historical/npsg_chapter_hap_jan2018.pdf [accessed 2021-05-08]
12. Hüske-Kraus D, Wilken M, Röhrig R. Measuring Alarm System Quality in Intensive Care Units. Zukunft der Pflege - Innovative Technologien für die Pflege 2018:89-93 [FREE Full text] [doi: 10.29007/5vpj]
13. Lorenz B, Peters J, Frey U. [How Much Alarm Can the Human Being Tolerate?]. Anasthesiol Intensivmed Notfallmed Schmerzther 2017 Jul 25;52(7-08):564-570. [doi: 10.1055/s-0042-118618] [Medline: 28743153]
14. Görges M, Markewitz B, Westenskow D. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg 2009 May;108(5):1546-1552. [doi: 10.1213/ane.0b013e31819bdfbb] [Medline: 19372334]
15. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010 Jan;19(1):28-34; quiz 35. [doi: 10.4037/ajcc2010651] [Medline: 20045845]
16. Allen JS, Hileman K, Ward A. Safety Solutions: Simple Solutions for improving Patient Safety In Cardiac Monitoring - Eight Critical Elements to Monitor Alarm Competency. AAMI Foundation Healthcare Technology Safety Institute. 2012. URL: http://s3.amazonaws.com/rdcms-aami/files/production/public/FileDownloads/HTSI/Alarm_Competency%20_White_Paper.pdf [accessed 2021-05-08]
17. Wilken M, Hüske-Kraus D, Röhrig R. Alarm Fatigue: Using Alarm Data from a Patient Data Monitoring System on an Intensive Care Unit to Improve the Alarm Management. Stud Health Technol Inform 2019 Sep 03;267:273-281. [doi: 10.3233/SHTI190838] [Medline: 31483282]
18. Cosper P, Zellinger M, Enebo A, Jacques S, Razzano L, Flack M. Improving Clinical Alarm Management: Guidance and Strategies. Biomed Instrum Technol 2017;51(2):109-115. [doi: 10.2345/0899-8205-51.2.109] [Medline: 28296432]
19. Welch J, Rayo M, Kanter B, Bagian T, Jacobs K, Shanawani H, et al. Framework for Alarm Management Process Maturity. Biomed Instrum Technol 2016;50(3):165-179. [doi: 10.2345/0899-8205-50.3.165] [Medline: 27164201]
20. Poncette AS, Wunderlich MM, Spies C, Heeren P, Vorderwülbecke G, Salgado E, et al. Resources for a "Do-it-Yourself Analysis" of the Patient Monitoring Alarm Data from Intensive Care Units. Zenodo 2021 Feb 24:1 [FREE Full text] [doi: 10.5281/zenodo.4560041]
21. The R Project for Statistical Computing. 2018. URL: https://www.R-project.org/ [accessed 2021-05-08]
22. dplyr: a grammar of data manipulation. URL: https://dplyr.tidyverse.org/ [accessed 2021-05-08]
23. tidyr: tidy messy data. URL: https://tidyr.tidyverse.org/ [accessed 2021-05-08]
24. Wickham H. stringr: modern, consistent string processing. The R Journal 2010;2(2):38. [doi: 10.32614/rj-2010-012]
25. Grolemund G, Wickham H. Dates and Times Made Easy with Lubridate. J. Stat. Soft 2011;40(3):25. [doi: 10.18637/jss.v040.i03]
26. Wickham H, Navarro D, Pedersen TL. ggplot2: elegant graphics for data analysis. New York, NY: Springer; 2016.
27. Li A, Gretzinger D. Applying an Evidence-based Approach to Managing Alarm Safety: A University Health Network Case Study. 2015 Presented at: World Congress on Medical Physics and Biomedical Engineering; June 7-12, 2015; Toronto, Canada. [doi: 10.1007/978-3-319-19387-8_374]
28. Gebrauchsanweisung IntelliVue Patientenmonitor MX600/MX700/MX800 Rev. Philips Healthcare. 2012. URL: https://www.usa.philips.com/healthcare/product/HC865241/intellivue-mx700-patient-monitor [accessed 2021-05-08]
29. Cvach M, Kitchens M, Smith K, Harris P, Flack M. Customizing Alarm Limits Based on Specific Needs of Patients. Biomed Instrum Technol 2017;51(3):227-234. [doi: 10.2345/0899-8205-51.3.227] [Medline: 28530858]
30. Fujita LY, Choi SY. Customizing Physiologic Alarms in the Emergency Department: A Regression Discontinuity, Quality Improvement Study. J Emerg Nurs 2020 Mar;46(2):188-198.e2 [FREE Full text] [doi: 10.1016/j.jen.2019.10.017] [Medline: 31864768]
31. Weiss B, Lütz A, Spies C. Delirium. In: O'Donnell J, Nácul F, editors. Surgical Intensive Care Medicine. Cham: Springer International Publishing; 2016:259-267.
32. Baillargeon E. Alarm Fatigue: A Risk Assessment. Rhode Island College. 2013 May 01. URL: https://digitalcommons.ric.edu/cgi/viewcontent.cgi?article=1218&context=etd [accessed 2021-05-08]
33. Deb S, Claudio D. Alarm fatigue and its influence on staff performance. IIE Transactions on Healthcare Systems Engineering 2015 Jul 29;5(3):183-196. [doi: 10.1080/19488300.2015.1062065]

XSL•FO
RenderX

34. Torabizadeh C, Yousefinya A, Zand F, Rakhshan M, Fararooei M. A nurses' alarm fatigue questionnaire: development and psychometric properties. J Clin Monit Comput 2017 Dec 15;31(6):1305-1312. [doi: 10.1007/s10877-016-9958-x] [Medline: 27848141]

Abbreviations

DIY: do-it-yourself
ECG: electrocardiogram
ICU: intensive care unit
IBP: invasive blood pressure
NIBP: noninvasive blood pressure
SpO_2: oxygen saturation

Edited by G Eysenbach; submitted 16.12.20; peer-reviewed by M Sayed, K Braune; comments to author 11.01.21; revised version received 11.03.21; accepted 02.04.21; published 28.05.21.

Please cite as:
Poncette AS, Wunderlich MM, Spies C, Heeren P, Vorderwülbecke G, Salgado E, Kastrup M, Feufel MA, Balzer F
Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions
J Med Internet Res 2021;23(5):e26494
URL: https://www.jmir.org/2021/5/e26494
doi:10.2196/26494
PMID:34047701

©Akira-Sebastian Poncette, Maximilian Markus Wunderlich, Claudia Spies, Patrick Heeren, Gerald Vorderwülbecke, Eduardo Salgado, Marc Kastrup, Markus A Feufel, Felix Balzer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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

Integrating Option Grid Patient Decision Aids in the Epic Electronic Health Record: Case Study at 5 Health Systems

Peter Scalia[1], MSc, PhD; Farhan Ahmad[2], BS, MBA; Danielle Schubbe[1], BA; Rachel Forcino[1], MSc, PhD; Marie-Anne Durand[1], PhD; Paul James Barr[1], MSc, PhD; Glyn Elwyn[1], PhD, MD

[1]Dartmouth College, Lebanon, NH, United States

[2]EBSCO Health, Boston, MA, United States

Corresponding Author:
Glyn Elwyn, PhD, MD
Dartmouth College
1 Medical Center Drive
Lebanon, NH, 03766
United States
Phone: 1 603 653 0806
Email: glynelwyn@gmail.com

Abstract

Background: Some researchers argue that the successful implementation of patient decision aids (PDAs) into clinical workflows depends on their integration into electronic health records (EHRs). Anecdotally, we know that EHR integration is a complex and time-consuming task; yet, the process has not been examined in detail. As part of an implementation project, we examined the work involved in integrating an encounter PDA for symptomatic uterine fibroids into Epic EHR systems.

Objective: This study aims to identify the steps and time required to integrate a PDA into the Epic EHR system and examine facilitators and barriers to the integration effort.

Methods: We conducted a case study at 5 academic medical centers in the United States. A clinical champion at each institution liaised with their Epic EHR team to initiate the integration of the uterine fibroid Option Grid PDAs into clinician-facing menus. We scheduled regular meetings with the Epic software analysts and an expert Epic technologist to discuss how best to integrate the tools into Epic for use by clinicians with patients. The meetings were then recorded and transcribed. Two researchers independently coded the transcripts and field notes before categorizing the codes and conducting a thematic analysis to identify the facilitators and barriers to EHR integration. The steps were reviewed and edited by an Epic technologist to ensure their accuracy.

Results: Integrating the uterine fibroid Option Grid PDA into clinician-facing menus required an 18-month timeline and a 6-step process, as follows: task priority negotiation with Epic software teams, security risk assessment, technical review, Epic configuration; troubleshooting, and launch. The key facilitators of the process were the clinical champions who advocated for integration at the institutional level and the presence of an experienced technologist who guided Epic software analysts during the build. Another facilitator was the use of an emerging industry standard app platform (Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources) as a means of integrating the Option Grid into existing systems. This standard platform enabled clinicians to access the tools by using single sign-on credentials and prevented protected health information from leaving the EHR. Key barriers were the lack of control over the Option Grid product developed by EBSCO (Elton B Stephens Company) Health; the periodic Epic upgrades that can result in a pause on new software configurations; and the unforeseen software problems with Option Grid (ie, inability to print the PDA), which delayed the launch of the PDA.

Conclusions: The integration of PDAs into the Epic EHR system requires a 6-step process and an 18-month timeline. The process required support and prioritization from a clinical champion, guidance from an experienced technologist, and a willing EHR software developer team.

(J Med Internet Res 2021;23(5):e22766) doi:10.2196/22766

KEYWORDS

shared decision making; patient decision aids; electronic health record; implementation; HL7 SMART on FHIR

XSL•FO
RenderX

Introduction

Background

Researchers have argued that the successful implementation of patient decision aids (PDAs) into clinical workflows depends on their integration into electronic health records (EHRs) [1,2]. The task of integrating third-party tools into EHRs is complex [3]. Security concerns dominate the challenge, as institutions have become reliant on EHRs to manage key operational workflows [3]. Third-party software that brings external connections and URL links to EHRs is subject to extensive scrutiny [4]. Updates to either the EHR or linked third-party products are perennial concerns, given the cost of downtime or system failure.

Many software vendors provide system-wide EHR software (eg, Epic, Cerner, or Allscripts) [5]. However, there are major differences between the same EHR product when installed at different health care institutions: this is because they are tailored to organizational and clinical preferences and integrated with other ancillary software [6]. Clinicians also differ in how and when EHRs are used with patients [7]. These uses can include showing images or test results and sending health information via the patient portal [8]. Although some common integration processes can be identified, solutions cannot be replicated from one institutional setting to others and tailoring is always required.

PDAs provide evidence-based information in a comparative format to facilitate shared decision-making, in which patients and clinicians are supported when making informed decisions together [9]. PDAs serve as catalysts to engage patients in decision-making processes and can be used before, during, and after clinical encounters [10,11]. Recent systematic reviews and meta-analyses have shown that PDAs increase knowledge about options and reduce decisional conflict, thereby helping patients make decisions that align with their preferences [12,13]. Despite improving a range of outcomes, their implementation in the clinical workflow remains a challenge [14,15]. Given the widespread adoption of EHRs and clinicians' reliance on them [16], many have presumed that integrating PDAs into EHRs will lead to their increased use in clinical practice [2,17]. However, this presumption has not yet been tested at scale.

Studies that have evaluated the integration of PDAs in EHRs have focused on measuring their use by clinicians [1,18-22], measuring their impact on patient outcomes [23-25], or user testing the tool to improve the navigation and design in the EHR system [26-29]. The integration of 2 PDAs, namely the Statin Choice and Diabetes Medication Choice tools, in the EHR at the Mayo Clinic led to their increased use [18,20,21]. Coylewright et al [1] also demonstrated an increased use and observed that adoption rates of an EHR-based *HealthDecision* tool steadily increased over an 8-year period, with a *high rate of sustained implementation after the fifth use*.

Nevertheless, we were only able to identify a few examples of PDAs being integrated into EHR systems [1,18-29]. Anecdotally, researchers and practitioners recognize that embedding PDAs in EHRs is a complex and time-consuming process; however, we could not identify the literature that described the required processes. Therefore, we lack an understanding of *how* best to integrate these tools into EHRs, and the steps required, especially given the recent development of new interoperability standards [30]. An opportunity arose to address this research gap as part of a project to implement the uterine fibroids Option Grid PDA at 5 health care institutions in the United States (Uterine Fibroids Options [UPFRONT] study) [31].

Objectives

The aims of this work are to (1) identify the steps and the time required to integrate an Option Grid PDA into the Epic EHR system and (2) examine facilitators and barriers to the integration effort. We hypothesize that some institutions will successfully integrate the Option Grid PDA into Epic as part of a multistep, time-intensive process.

Methods

Design

As part of a Patient-Centered Outcomes Research Institute (PCORI)–funded, stepped-wedge implementation trial, we asked each participating institution to integrate Option Grid PDAs into their Epic EHR systems. Despite the stepped-wedge design, we began the integration effort at all institutions almost immediately upon receiving funding to provide ample opportunity to complete the process ahead of the active implementation phase of the broader trial, which is when clinicians would be expected to use Option Grid with their patients. Successful integration was defined as the completion of changes to the Epic system that allowed clinicians to easily access an external website that provides access to both interactive and PDF versions of the uterine fibroid Option Grid PDAs [31]. Facilitators are key elements or factors that enable a successful integration. To examine the processes required, we adopted an exploratory case study design and collected data by recording relevant meetings and taking field notes [32]. We analyzed our conversations with the clinical champions, their research teams, and Epic software analysts from various departments (such as compliance, risk management, and information security) at each institution. Our case study was reported using the checklist by Rodgers et al (Multimedia Appendix 1) [33]. The Dartmouth College Committee for the Protection of Human Subjects (approval number: STUDY00031464) granted ethical approval for our study.

Settings

The implementation study took place in the following institutions: (1) Dartmouth-Hitchcock Medical Center in Lebanon, New Hampshire; (2) Barnes-Jewish Hospital in St. Louis, Missouri; (3) Montefiore Medical Center in Bronx, New York; (4) Brigham and Women's Hospital in Boston, Massachusetts; and (5) Mayo Clinic in Rochester, Minnesota. Each of these institutions had installed the Epic EHR product at different times. Our case study is based on our efforts to integrate the Option Grid into the Epic EHR system at these 5 institutions. Table 1 provides a brief description of each institution's Epic experience, expertise, and infrastructure.

XSL•FO
RenderX

Table 1. Description of each institution's Epic experience, expertise, and infrastructure.

Institution	Date of Epic adoption	Number of Epic software analysts[a]	Any previous experience with third-party software integration?	Does the clinical champion have experience with using patient decision aids?
Dartmouth-Hitchcock Medical Center	April 2011	170	Yes	Yes
Barnes-Jewish Hospital	June 2017	150	Yes	No
Montefiore Medical Center	April 2015 to June 2016	200	Yes	No
Brigham and Women's Hospital	June 2015	Unknown	Yes	Yes
Mayo Clinic	May 2018	300	Yes	No

[a]Estimated number.

Option Grid PDA

The Option Grid PDA for symptomatic uterine fibroids is part of a suite of tools developed and updated by EBSCO (Elton B Stephens Company) Health, a commercial entity that provides clinical decision support for health care organizations [34]. Our collaboration with EBSCO Health for developing and maintaining the uterine fibroids PDA came at no cost to the research effort. The uterine fibroid tool compares 7 treatment options: (1) watch and wait, (2) medicine with hormones, (3) medicine without hormones, (4) embolization, (5) endometrial ablation, (6) myomectomy, and (7) hysterectomy. The tool is available in English and Spanish for the following 2 formats: text-only and text accompanied by pictures (Picture Option Grid). For the study duration, each participating institution was granted access to the entire suite of 30 Option Grid tools.

When clinicians click the Option Grid button, they are first presented with the entire suite of PDAs. Once clinicians select the uterine fibroids Option Grid, they have the opportunity to select as many as 7 treatment options that are relevant to a particular patient. Once the PDA is generated, the clinician can use 3 features to document the options discussed in the encounter: print a PDF version, copy and paste a script that includes the options selected, or send a *permalink* to the patient so that they can view the Option Grid at their own convenience. The script is an optional feature that enables clinicians to document the use of the PDA and the options discussed in the EHR. The clinician must either use this feature or write a note in the EHR regarding the conversation that occurred with the PDA, as Option Grid does not exchange any information with the EHR. Figure 1 shows an example of an Option Grid PDA for symptomatic uterine fibroids.

Figure 1. Snapshot of the text version of the uterine fibroid Option Grid patient decision aid.



Recruitment

The 5 institutions were selected to participate in our implementation trial because of their inherent diversity, their interest in implementing PDAs, and, in some cases, their experience of practicing shared decision-making. Institutions (from inner city Bronx to rural New Hampshire) treat an ethnically diverse patient population across both urban and rural settings. At each of the 5 institutions, an obstetrics and gynecology specialist was recruited for the role of clinical

XSL•FO
RenderX

champion (site principal investigator) for their interest in using PDAs to improve health care delivery. The clinical champions contacted the Epic software analysts and explained the importance of integrating the uterine fibroids Option Grid PDA into the EHR. The clinical champion also introduced software analysts to the UPFRONT study team, which included an experienced Epic technologist employed at EBSCO Health (FA).

Data Collection

Videoconference meetings were scheduled between the UPFRONT team and the Epic software analysts at each institution. Following the introductory meeting, the Epic software analysts dictated the frequency of the meetings based on their needs for assistance or clarification from the Epic technologist. The software analysts reached out to our study team to schedule meetings. The meetings were a time to collect the work summaries that had been collected by the software analysts, identify any barriers they were facing, and discuss solutions to overcome those barriers. Each meeting was audio-recorded. The transcribed audio recordings and field notes from the meetings provided the data for analysis.

Consent

The Dartmouth College Committee for the Protection of Human Subjects waived the requirement for the written documentation of informed consent. Standard information regarding the study was provided to the participants. The study team sought verbal consent from the audio recording at the start of videoconference meetings. If verbal consent was not granted by one or more members, then the study team took notes of meeting discussions.

Analysis

For aim 1, 2 researchers (PS and DS) reviewed meeting transcripts and field notes and documented the steps taken to integrate the Option Grid into Epic. Our description of the steps was reviewed and edited by the Epic technologist (FA) and modifications were made if required. To determine the amount of time required for the integration, we counted the number of months from the original email to the clinical champions to gauge their interest in a possible integration effort to the moment the Option Grid was launched in the site's Epic environment. For aim 2, we conducted an inductive thematic analysis of the transcripts and field notes. Two researchers (PS and DS) independently coded a sample of transcript pages and then discussed and agreed on a codebook. The finalized codebook (Multimedia Appendix 2) was applied to all the data to highlight the facilitators and barriers to the integration effort. Codes were grouped into different code categories, which were revised and discussed by PS and DS to determine the themes. Coding disagreements were resolved by a third researcher (GE).

Results

Overview

A total of 27 meetings were held. These included 25 videoconferences (Zoom, Webex [Cisco], or Skype [Microsoft Corporation]) and 2 telephone meetings. Of the 27 meetings, we were able to record 22 (81%) meetings and collated notes for the remaining 5 (19%). The bidirectional exchange of information between the Epic software analysts and the study team (including the Epic technologist) during the meetings at each institution yielded a total of 183 transcripts or field note pages (total word count: 91,336; see Table 2 for details).

XSL•FO
RenderX

Table 2. Details of the 27 meetings that informed the steps to integrate the patient decision aid into the electronic health record and the facilitators and barriers to the integration effort.

Institution and meeting date	Platform	Recorded? (yes or no)	Meeting duration (min:sec)	Meeting personnel
Dartmouth-Hitchcock Medical Center				
January 9, 2019	Zoom	Yes	40:03	3 UPFRONT[a] study members, 4 EBSCO[b] Health members, and 2 clinical informatics members and clinicians
February 6, 2019	Phone	Yes	44:50	2 UPFRONT study members, 1 clinical champion, 1 Epic technologist, 1 EBSCO Health member, 1 clinical informatics member and clinician, and 2 clinicians
July 25, 2019	Webex	Yes	23:39	3 UPFRONT study members and 1 Epic technologist
August 9, 2019	Zoom	Yes	5:49	2 UPFRONT study members, 1 Epic technologist, and 1 clinical informatics member and clinician
Barnes-Jewish Hospital				
February 4, 2019	Zoom	Yes	50:24	3 UPFRONT study members, 1 clinical champion, 1 Epic technologist, 2 EBSCO Health members, and 1 research assistant
May 23, 2019	Zoom	Yes	27:25	3 UPFRONT study members, 1 Epic technologist, and 1 Epic software analyst
June 6, 2019	Zoom	Yes	35:07	3 UPFRONT study members, 2 Epic software analysts, and 1 Epic technologist
February 3, 2020	Webex	Yes	MDNR[c]	1 UPFRONT study member, 1 Epic software analyst, 2 clinicians, and 18 ambulatory operations members
February 11, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, 1 Epic software analyst, and 1 interfaces team member
March 3, 2020	Webex	Yes	27:35	1 UPFRONT study member, 1 Epic technologist, and 1 Epic software analyst
Montefiore Medical Center				
January 22, 2019	Zoom	Yes	7:45	2 clinical champions, 1 UPFRONT study member, 1 Epic software analyst, and 3 research assistants
January 30, 2019	Zoom	Yes	27:38	2 UPFRONT study members, 2 EBSCO Health members, and 4 Epic software analysts
January 14, 2020	Skype	No	__[d]	1 clinical champion, 2 UPFRONT study members, 1 Epic technologist, 2 Epic software analysts, and 2 research assistants
Brigham and Women's Hospital				
January 25, 2019	Zoom	Yes	21:59	2 UPFRONT study members and 1 partners operations member
February 18, 2019	Zoom	Yes	33:28	2 UPFRONT study members and 2 Epic software analysts
March 27, 2019	Zoom	Yes	15:45	2 UPFRONT study members and 2 Epic software analysts
May 22, 2019	Zoom	Yes	37:08	3 UPFRONT study members, 1 Epic technologist, 5 Epic software analysts, and 1 research assistant
July 29, 2019	Zoom	Yes	18:42	2 UPFRONT study members, 1 Epic technologist, 1 partners operations member, and 2 research assistants
January 9, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
March 30, 2020	Webex	Yes	MDNR	1 clinical champion, 1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
April 24, 2020	Phone	No	—	1 UPFRONT study member, 1 Epic technologist, 1 research assistant, and 2 information security analysts
May 28, 2020	Webex	No	—	1 Clinical champion, 1 Epic technologist, 2 Epic software analysts, and 1 research assistant
June 5, 2020	Webex	No	—	1 UPFRONT study member, 1 Epic technologist, 2 Epic software analysts, and 1 EBSCO Health member

XSL•FO
RenderX

Institution and meeting date	Platform	Recorded? (yes or no)	Meeting duration (min:sec)	Meeting personnel
June 23, 2020	Webex	No	—	1 UPFRONT study member, 1 Epic technologist, and 2 Epic software analysts
Mayo Clinic				
January 23, 2019	Webex	Yes	47:59	1 Clinical champion, 1 UPFRONT study member, and 3 Epic software analysts
July 24, 2019	Zoom	Yes	47:33	1 clinical champion, 3 UPFRONT study members, 1 Epic technologist, 3 Epic software analysts, and 1 research assistant
January 29, 2020	Webex	Yes	MDNR	1 UPFRONT study member, 1 Epic technologist, and 3 Epic software analysts

[a]UPFRONT: Uterine Fibroids Options.

[b]EBSCO: Elton B Stephens Company.

[c]MDNR: meeting duration not recorded.

[d]Not available. Participants did not agree to be recorded.

Aim 1: The Steps Taken to Integrate the PDA Into Epic

We were able to describe the process of integrating the Option Grid PDAs into Epic by identifying 6 common process steps across the institutions (Textboxes 1 and 2).

The timeline for completing the 6 steps varied across institutions, but overall, up to 18 months (January 2019 to June 2020) was required to integrate the Option Grid PDAs into the Epic EHR system. The timeline began in January 2019, which was when all clinical champions received an email to gauge interest in integrating Option Grid into a clinician-facing menu in their Epic system. Although work began to place a button in an agreed location nominated by clinicians at each institution, a policy decision was made by EBSCO Health in August 2019 to use Substitutable Medical Applications and Reusable Technologies (SMART) and Fast Healthcare Interoperability Resources (FHIR) standards and list the application on Epic's App Orchard to simplify the setup and maintenance of the tool throughout the study period. Owing to the COVID-19 pandemic, integration efforts at 2 institutions were paused to redirect resources to other more pressing initiatives. One institution was able to resume the effort, whereas the other institution had competing priorities and was forced to furlough some personnel involved in the integration process. Thus, we were only able to complete the integration of Option Grid into Epic at 4 of the 5 sites. Figure 2 details the integration timeline following the policy decision, beginning with the security risk assessment (Step 2) and ending with the Option Grid launch (Step 6).

XSL•FO
RenderX

Textbox 1. Steps to integrate the Option Grid patient decision aids into the electronic health record.

Step 1: Negotiating task priority with Epic software teams

- The clinical champions at each site requested the changes into their Epic systems, establishing the clinical benefit of providing easy access to the Option Grid website in a clinician-facing menu location. Epic teams had to reprioritize their tasks, given existing work schedules. In one setting, agreement to reprioritize required negotiation and financial support

Step 2: Security risk assessment

- Each institution had different security risk assessment processes, with each requiring departmental approval. Typically, 3 levels of security checks were required, related to the changes in outpatient processes or menus, information flows and dependencies, and communication with third-party tools. Epic has different modules (ie, outpatient, inpatient, and research) and each has its own operations and approval groups. The Ambulatory Operations group is responsible for reviewing requests related to general outpatient workflows (Barnes-Jewish Hospital and Montefiore Medical Center). In the case of Barnes-Jewish Hospital, the integration had to also be approved by the Interfaces Operations group, which manages requests related to any information that gets moved in and out of Epic via interfaces, and the Infrastructure team, which manages the App Orchard. The Mayo Clinic also had 3 levels of security: (1) Security, Privacy, Architecture and Data Assurance reviews new technology that will be integrated into Epic; (2) Clinical Decision Support reviews electronic health record (EHR) change requests that have endorsement from a clinical or practice committee and includes some form of clinical decision support (ie, patient decision aids [PDAs]); and (3) the Obstetrics and Gynecology specialty group, which reviews any new process, procedure, or app that will be integrated into Epic from the clinical perspective. The 4 levels of security at Brigham and Women's Hospital include (1) the clinical vetting that reviews the study context with 2 clinicians; (2) the technical feasibility of the project evaluated by an Epic team leader, (3) technical assessment (see Step 3 for details); and (4) the security risk assessment that represents an internal process to review the application being integrated into Epic. Security review was considered unnecessary at Dartmouth-Hitchcock because the Option Grid tools were merged with an existing decision support product—HealthDecision. Analysts were particularly focused on data exchange requirements between the Option Grid app and Epic. The lack of protected health information (PHI) transfer was important. Using Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources to allow approval for the synchronization of the private Option Grid App Orchard app with the institution's Epic environment (Step 4 for details) standardized the process across all 5 sites

Step 3: Technical review

- The Epic software analysts determined the number of personnel and time required for the overall software changes. Once the level of effort, allocation of tasks, and timelines were established, the software analysts were ready for the build

Step 4: Epic configuration

- Before commencing the build, all clinical champions indicated their preference for placing the Option Grid button in the patient's chart under the More menu in their toolbar at the top. Some clinical champions preferred that the button be accessible to the entire institution, whereas others preferred a restricted access to their obstetrics and gynecology department. Our study team's Epic technologist then developed and shared a build guide with each institution. The guide outlined the study objectives, the Epic-specific configurations required, and described how the Option Grid would not require access to PHI. The Epic software analysts configured access to the Option Grid PDA using Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources (see Textbox 2 for further details). The guide contained 4 steps:
    - Accessing the app at the Epic App Orchard
    - Enabling synchronization of the Option Grid PDA on the App Orchard with the institution's Epic environment
    - Establishing a test environment before launch in a production environment
    - Requesting whitelisting of the Option Grid domains so that the app could be accessed within Epic's EHR menus, tasks, and options. Epic uses the term Hyperspace to describe this view of the software. Launching Option Grid within Hyperspace allows clinicians to have an easy access to the PDAs, without having to navigate to an external website

Step 5: Troubleshooting

- The software analysts ensured that menu locations, access requests, and user identification were all functioning as planned. This step represented a final check to ensure that all the Option Grid features were operational

Step 6: Launch

- After troubleshooting, the new configuration was migrated to the production environment and the Option Grid was launched. This means that clinicians could access the Option Grid button and be directed to the Option Grid website where they can generate a PDA for their patients

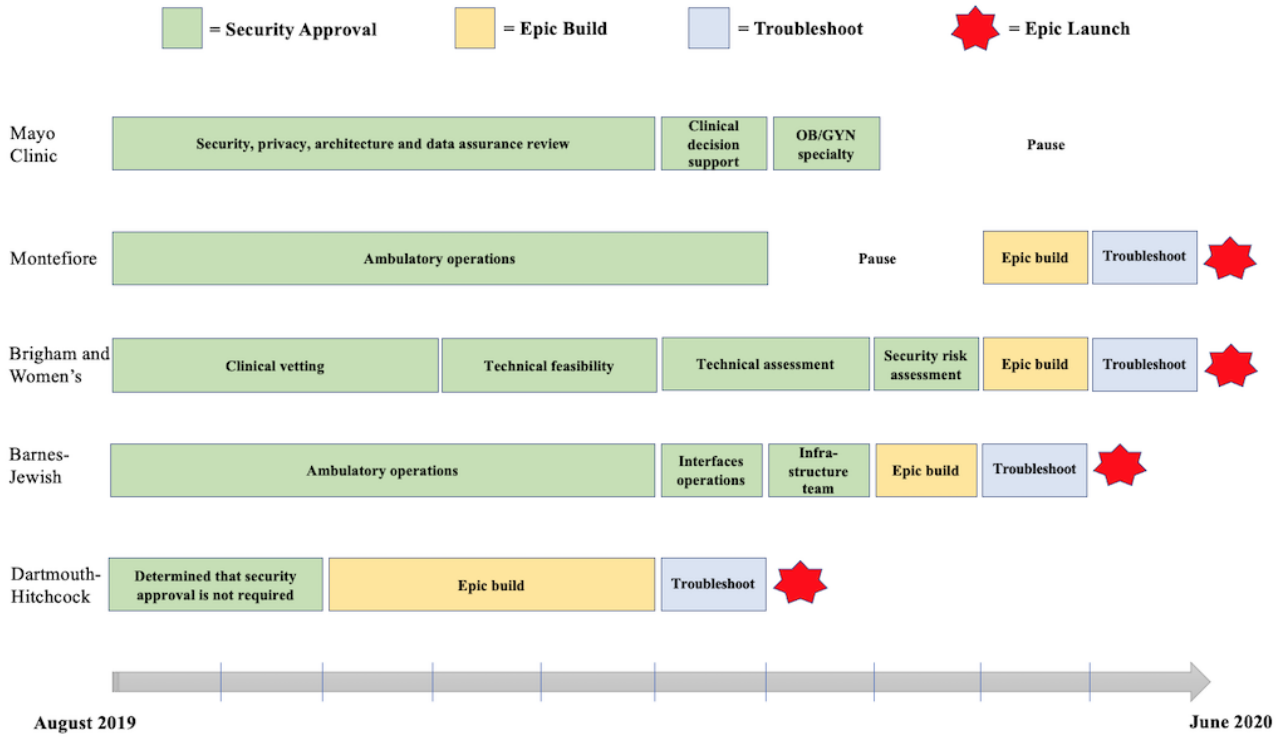
XSL•FO
RenderX

Textbox 2. Health Level 7 Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources authentication.

SMART on FHIR

- Substitutable Medical Applications and Reusable Technologies (SMART) is an "open, standards-based platform that enables innovators to create applications that seamlessly and securely run across the healthcare system" [35]. Health Level 7 (HL7) is an "industry organization that develops standards for the exchange, integration, sharing and retrieval of EHR information" [36]. HL7 adopted the OAuth2-based SMART App Launch framework as a core interoperability standard. HL7 also developed the Fast Healthcare Interoperability Resources (FHIR) standard to ensure *interoperability, extensibility, and speed* while searching for information across clinical applications [37]. SMART, along with FHIR—collectively referred to as SMART on FHIR—connects third-party apps to Epic, enabling them to reliably and securely launch in Epic's Hyperspace desktop client [38]. We used SMART on FHIR authentication for the following 3 reasons:
  - Improved analytics: allows the tracking of app use so that we can determine the number of eligible patients who received the uterine fibroid Option Grid patient decision aid (PDA)
  - Improved control for data access: SMART on FHIR allows a better control of the information shared with third-party apps
  - Uses existing authorizations: using FHIR allows clinician access to the Option Grid PDAs in their existing user interface (Epic's Hyperspace)
- To leverage SMART on FHIR, Epic requires apps to participate in its App Orchard app store. For the UPFRONT (Uterine Fibroids Options) study, EBSCO (Elton B Stephens Company) chose to list Option Grid as a private app on the Epic App Orchard and allowed access by the 5 participating institutions. This also simplified the setup and maintenance of the PDA app during the study period.

Figure 2. Timeline of the electronic health record integration of the Option Grid patient decision aid. OB/GYN: obstetrics and gynecology.



Aim 2: Thematic Analysis of Facilitators and Barriers to EHR Integration

We identified 4 integration facilitators (Textbox 3):

Textbox 3. Facilitators of electronic health record integration.

Facilitators

1. Clinical advocacy: presence of a clinical champion at each institution
2. Electronic health record expertise: presence of an Epic technologist with experience in building apps in Epic
3. Standardization of process: use of Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources standards
4. Avoidance of protected health information (PHI) data transfer: no exchange of PHI between Option Grid and the institution's local Epic environment

XSL•FO
RenderX

Clinical Advocacy

The clinical champions provided the required professional arguments for integration and served as a gateway to the Epic software analysts and lobbied for prioritization of the task. For instance, the principal investigator at Montefiore Medical Center advocated for the Epic software analysts to prioritize the integration effort:

> *We are getting it prioritized by the Epic work group with a goal for the summer.* [Principal investigator at the Montefiore Medical Center]

EHR Expertise

Having an EHR expert on our study team with a direct experience of Epic EHR environments and of the app's requirements at EBSCO was a major facilitator. He developed a *build guide*, which facilitated the integration of Option Grid PDAs in Epic at the 5 settings. Throughout the meetings, the EHR expert supported the software analysts by answering questions, clarifying points of confusion, and providing solutions for any technical issues that arose throughout the process. The following is an example of an exchange between the Epic expert and a software analyst at Barnes-Jewish Hospital during the troubleshooting step of the integration process:

> *The environment is listed in our App Orchard listing and has the test client ID. Can we test? [launch URL provided] The tokens in the OAuth context values are the same.* [Epic technologist]
>
> *Thank you—I updated the URL item in the record and tested it again. That seemed to work! I am able to open the Uterine Fibroids Option Grid, copy and paste information, open a PDF, and print the PDF.* [Barnes-Jewish hospital software analyst]

Standardization of Process

First, the use of SMART on FHIR standards provided a helpful and officially sanctioned way to better control the information being shared with a third party such as the Option Grid PDA. As described in Textbox 2, the process provided clinicians with single sign-on credentials and enabled us to track the number of times each clinician generated or accessed a uterine fibroid PDA. The Epic technologist also indicated that Epic recommends SMART on FHIR over alternative launch methods such as active guidelines and other URL-based methods:

> *The reason we went with SMART authentication is because Epic specifically advised us to do so for newer implementations.* [Epic technologist]

Avoidance of Protected Health Information Data Transfer

Placing the Option Grid PDA as a private app in Epic's App Orchard restricted Option Grid from retrieving protected health information (PHI). Option Grid instructed Epic to block certain features (ie, incoming application programming interface), so their organization would be unable to attain PHI. This arrangement eased the security concerns at each institution. For example, a software analyst at the Mayo Clinic informed us that it would be easier to obtain approval from the security team if PHI was not leaving their Epic environment:

> *When we bring in new technology it has to go to Security, Privacy, Architecture and Data Assurance (SPAD) review, but if there is no PHI it will get blessed faster.* [Mayo software analyst]

We identified 3 themes that represented barriers to the integration effort (Textbox 4).

Textbox 4. Barriers of electronic health record integration.

Barriers
1. Commercial third-party autonomy: lack of control over the Option Grid patient decision aid product owned by EBSCO Health
2. Electronic health record updates and maintenance: periodic Epic upgrades causing some delays and functionality issues
3. Unforeseen software problems: found while troubleshooting the app leading to minor delays to launch

Commercial Third-Party Autonomy

Although our collaboration with EBSCO was an overall facilitator, it also presented a barrier. EBSCO owns the Option Grid product, and as researchers, we have no influence on product development. Future product development may require additional integration efforts during the life of the study and remain a risk area. For instance, the policy decision in August 2019 to change the integration strategy forced software analysts to adapt and use a different configuration than originally planned. The following is a part of the study team's communication to each institution, explaining the shift in integration strategy:

> *EBSCO has been working on enhancing Option Grid and its integration with Epic. This process will utilize SMART and FHIR standards. As a result, we are requesting to delay the integration build in your Epic environment.* [UPFRONT study team]

EHR Updates and Maintenance

Some institutions have regular Epic upgrades and either refrain from integrating apps during those upgrades or impose a freeze on all new software configurations. The software analysts did their best to plan ahead and sidestep this barrier at the final step. The following are 2 quotes that highlight reluctance to launch the button during an upgrade:

> *Some customers have code freezes that could last more than a month when they have Epic upgrades.* [EHR technologist]
>
> *We generally don't push out new applications during an Epic upgrade.* [Brigham software analyst]

XSL•FO
RenderX

Unforeseen Software Problems

We experienced unforeseen software-related problems during the troubleshooting steps. In some settings, clinicians were unable to print PDF versions of the tool or copy and paste the script documenting the uterine fibroid options discussed in the patient's record. These software issues were conveyed to our EHR technologist:

> *We confirmed that the copy/paste does not work when launching the application.* [Dartmouth software analyst]
>
> *The PDF button did not work. I was still able to navigate elsewhere on the page, but the PDF button was non-responsive.* [Brigham software analyst]

These issues were quickly resolved through collaboration between the software analysts and the Epic technologist and did not significantly delay the launch of the Option Grid in the site's Epic environments.

Discussion

Principal Findings

The integration of Option Grid PDAs into an EHR such as Epic requires clinical advocacy, a standardized process that avoids using PHI, and expertise to guide the process. Without the support of a clinical champion in each setting, we would not have been able to initiate the process of PDA integration. At the core of the work are issues of security and reassuring the organization that data transfers will not breach security protocols. SMART on FHIR addresses the data security requirements by allowing for a better control of the information being shared with a third party such as Option Grid. The availability of an EHR expert on our study team provided the necessary guidance and reassurance to the existing Epic teams. With all these components and facilitators present, the integration process took up to 18 months to achieve. Barriers were the lack of control over the Option Grid product, EHR updates and maintenance, and unforeseen software problems that caused delays and functionality issues.

Strengths and Limitations

Our use of a case study method to elicit a real-world, in-depth understanding of the steps required to integrate PDAs into Epic, and the associated barriers and facilitators, is novel and provides new insights. However, this study has limitations. First, we were only able to examine the Epic EHR system, so we do not know if our description of the integration process steps applies to other systems. Second, the Option Grid PDAs do not require the use of PHI data. Some PDAs being developed require the exchange of PHI, which we suspect would prolong integration timelines at many institutions. Third, all the institutions were large academic medical centers with expertise in configuring the Epic EHR system, so we do not know if our findings are applicable to smaller clinical practices that lack such capability. Fourth, we did not conduct a thematic analysis to address our first aim. However, despite the absence of a thematic analysis, we feel like a review of meeting transcripts, with oversight from an Epic technologist, represents the appropriate method to determine the steps to integrate the Option Grid PDA into Epic. Furthermore, it is not known whether EBSCO Health would provide an Epic technologist to other customers or organizations aiming to integrate Option Grid in their Epic. Finally, because of the COVID-19 pandemic, we were only able to integrate the Option Grid PDAs at 4 of the 5 institutions. One institution had to redirect resources to other initiatives and were faced with staffing limitations, which led to a pause in the integration effort. We did not include this as a barrier, considering that under normal circumstances, the institution would be positioned to complete the integration of Option Grid in their Epic.

Results in Context

To the best of our knowledge, this is the first study to describe SMART on FHIR standards to integrate third-party PDAs, such as Option Grid, into an EHR system. Our results address an important gap outlined by a recent feasibility study that integrated a third-party prostate cancer screening PDA app into the EHR [3]. The authors of that study recognized the potential of SMART on FHIR to standardize secure data exchange and enable integration across a variety of EHRs [3]. However, research has focused on the interoperability of FHIR standards. For instance, a recent review showed how FHIR moved clinical information (medical images and quality metrics) found on different platforms in the EHR into a single platform to streamline the workflow of radiologists [37]. Similarly, another system has used FHIR standards to collect data from multiple sources in the EHR, automate analyses of laboratory test results, and generate easy-to-read reports for patients and their clinicians [39].

For aim 2, a key facilitator of Option Grid PDA integration into Epic systems was the presence of a clinical champion. This aligns with the results of an effort to integrate an EHR-based PDA in the emergency department for concussion and brain injury decisions [40]. They reported the critical need to engage clinicians and other information technology stakeholders [40]. In our case, the clinical champion served as an intermediary between the study team and the Epic software analysts, facilitated prioritization, and identified the EHR menu button location to ensure visibility. Clinicians' input in the integration process is reported to potentially cause the sustained use of the tools in practice [41,42] and is key to an integration effort, regardless of the format or mode of delivery [14,43-45].

Implications

Integrating third-party software into EHR systems requires a clinical champion to advocate for the task at the institutional level and an EHR expert who can guide software analyst teams throughout the process. From a policy perspective, implementing SMART on FHIR-compatible servers, which has been done at Duke Medical Center, can improve interoperability and the seamless integration of patient-facing apps [46]. However, the technologies for standardizing the integration of various types of apps, such as Option Grid, do not necessarily mean that they will be used in clinical practice. Integration, though difficult to achieve, seems to be the first step to ensure that clinicians have access to such tools. Providing an integration guide for other organizations to follow and identifying the barriers and facilitators of the process will enable more tools to be integrated into workflows. However, access alone might not lead to use,

XSL•FO
RenderX

as observed by others [21]. Future work should include usability, acceptability, or health technology assessments to further evaluate how to trigger clinicians to access and use embedded PDAs [22].

Conclusions

Integrating the uterine fibroid Option Grid PDA into clinician-facing menus in Epic was an approximately 18-month process, facilitated by a clinical champion who lobbied for the prioritization of the effort at the institutional level, and an EHR expert who guided the Epic software analysts throughout the study. The use of Health Level 7 SMART on FHIR standardized the integration effort, provided clinicians with single sign-on credentials, and more importantly blocked the exchange of PHI between Epic and Option Grid PDAs. Whether integration leads to patient use remains an open question.

Acknowledgments

The authors would like to thank the clinical champions at each institution for spearheading the EHR integration effort and serving as the gateway to the Epic software analysts: Dr Erika Banks, Dr Raymond M Anchan, Dr Shannon K Laughlin-Tommaso, Dr Tessa Madden, and Dr Tina C Foster. They also appreciate the time and effort of every Epic software analyst they encountered in their journey. They want to thank Dr Jon Keevil at EBSCO for working with site personnel to synchronize the private Option Grid App Orchard app with their Epic environment. Research reported in this study was funded by the PCORI award (grant SDM-2017C2-8507). The statements presented in this study are solely the responsibility of the authors and do not necessarily represent the views of the PCORI, its Board of Governors, or the Methodology Committee.

Conflicts of Interest

GE has edited and published books that provide royalties on sales by publishers, including Shared Decision Making (Oxford University Press) and Groups (Radcliffe Press). GE's academic interests focus on shared decision making and coproduction. He owns copyright in measures of shared decision making and care integration, namely collaboRATE, integRATE (measure of care integration), consideRATE (patient experience of care in serious illness), coopeRATE (measure of goal setting), toleRATE (clinician attitude to shared decision making), Observer OPTION-5, and Observer OPTION-12 (observer measures of shared decision making). In the past, he provided consultancy for organizations, including (1) Emmi Solutions LLC (limited liability company) who developed patient decision support tools; (2) the National Quality Forum on the certification of decision support tools; (3) Washington State Health Department on the certification of decision support tools; and (4) SciMentum LLC, Amsterdam (workshops for shared decision making). He is the founder and director of &think LLC, which owns the registered trademark for Option Grid PDAs, and founder and director of SHARPNETWORK LLC, a provider of training for shared decision making. He provides advice in the domain of shared decision making and PDAs to (1) Access Community Health Network, Chicago (Adviser to Federally Qualified Medical Centers); (2) EBSCO Health for Option Grids PDAs (consultant); (3) Bind On Demand Health Insurance (consultant); (4) PatientWisdom Inc (Adviser); and (5) Abridge AI (artificial intelligence) Inc (Chief Clinical Research Scientist).

MAD is a consultant to Access Community Health Network. Together with GE, she developed the Option Grid PDAs, which are licensed to EBSCO Health. She receives consulting income from EBSCO Health and may receive royalties in the future.

FA is the Vice President of Clinical Decision Support Technology at EBSCO Health with existing and future applications on the Epic App Orchard and a previous Epic employee.

No other authors have conflicts of interest to declare.

Multimedia Appendix 1
The guidelines for the reporting of organizational case studies.
[DOCX File , 13 KB - jmir_v23i5e22766_app1.docx]

Multimedia Appendix 2
Code hierarchy used to identify the facilitators and barriers of the electronic health record integration effort.
[DOCX File , 20 KB - jmir_v23i5e22766_app2.docx]

References

1. Coylewright M, Keevil JG, Xu K, Dodge SE, Frosch D, Field ME. Pragmatic Study of Clinician Use of a Personalized Patient Decision Aid Integrated into the Electronic Health Record: An 8-Year Experience. Telemed J E Health 2020 May;26(5):597-602. [doi: 10.1089/tmj.2019.0112] [Medline: 31381477]
2. Lenert L, Dunlea R, Del FG, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
3. Day FC, Pourhomayoun M, Keeves D, Lees AF, Sarrafzadeh M, Bell D, et al. Feasibility study of an EHR-integrated mobile shared decision making application. Int J Med Inform 2019 Dec;124:24-30. [doi: 10.1016/j.ijmedinf.2019.01.008] [Medline: 30784423]

XSL•FO
RenderX

4. Tejero A, de LTI. Advances and current state of the security and privacy in electronic health records: survey from a social perspective. J Med Syst 2012 Oct;36(5):3019-3027. [doi: 10.1007/s10916-011-9779-x] [Medline: 21947856]
5. Ratwani RM, Savage E, Will A, Arnold R, Khairat S, Miller K, et al. A usability and safety analysis of electronic health records: a multi-center study. J Am Med Inform Assoc 2018 Sep 01;25(9):1197-1201. [doi: 10.1093/jamia/ocy088] [Medline: 29982549]
6. Ben-Assuli O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health Policy 2015 Mar;119(3):287-297. [doi: 10.1016/j.healthpol.2014.11.014] [Medline: 25483873]
7. McGinn T. Putting Meaning into Meaningful Use: A Roadmap to Successful Integration of Evidence at the Point of Care. JMIR Med Inform 2016;4(2):e16 [FREE Full text] [doi: 10.2196/medinform.4553] [Medline: 27199223]
8. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
9. Elwyn G, Lloyd A, Joseph-Williams N, Cording E, Thomson R, Durand M, et al. Option Grids: shared decision making made easier. Patient Educ Couns 2013 Feb;90(2):207-212. [doi: 10.1016/j.pec.2012.06.036] [Medline: 22854227]
10. Tsulukidze M, Grande SW, Gionfriddo MR. Assessing Option Grid® practicability and feasibility for facilitating shared decision making: An exploratory study. Patient Educ Couns 2015 Jul;98(7):871-877. [doi: 10.1016/j.pec.2015.03.013] [Medline: 25841546]
11. Seal RP, Kynaston J, Elwyn G, Smith PEM. Using an Option Grid in shared decision making. Pract Neurol 2014 Mar;14(1):54-56. [doi: 10.1136/practneurol-2013-000666] [Medline: 24151338]
12. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
13. Scalia P, Durand M, Berkowitz JL, Ramesh NP, Faber MJ, Kremer JAM, et al. The impact and utility of encounter patient decision aids: Systematic review, meta-analysis and narrative synthesis. Patient Educ Couns 2019 May;102(5):817-841. [doi: 10.1016/j.pec.2018.12.020] [Medline: 30612829]
14. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards AGK, Clay C, et al. "Many miles to go …": a systematic review of the implementation of patient decision support interventions into routine clinical practice. BMC Med Inform Decis Mak 2013;13 Suppl 2:S14 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S14] [Medline: 24625083]
15. Dobler CC, Sanchez M, Gionfriddo MR, Alvarez-Villalobos NA, Singh Ospina N, Spencer-Bonilla G, et al. Impact of decision aids used during clinical encounters on clinician outcomes and consultation length: a systematic review. BMJ Qual Saf 2019 Jun;28(6):499-510 [FREE Full text] [doi: 10.1136/bmjqs-2018-008022] [Medline: 30301874]
16. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
17. Ankolekar A, Dekker A, Fijten R, Berlanga A. The Benefits and Challenges of Using Patient Decision Aids to Support Shared Decision Making in Health Care. JCO Clin Cancer Inform 2018 Dec;2:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00013] [Medline: 30652607]
18. Inselman J, Branda M, Castaneda-Guarderas A, Gionfriddo MR, Zeballos-Palacios CL, Morris MM, et al. Uptake and Documentation of the Use of an Encounter Decision Aid in Usual Practice: A Retrospective Analysis of the Use of the Statin/Aspirin Choice Decision Aid. Med Decis Making 2016 May;36(4):557-561. [doi: 10.1177/0272989X15618175] [Medline: 26721257]
19. Jouni H, Haddad RA, Marroush TS, Brown S, Kruisselbrink TM, Austin EE, et al. Shared decision-making following disclosure of coronary heart disease genetic risk: results from a randomized clinical trial. J Investig Med 2017 Mar;65(3):681-688 [FREE Full text] [doi: 10.1136/jim-2016-000318] [Medline: 27993947]
20. Ye S, Leppin AL, Chan AY, Chang N, Moise N, Poghosyan L, et al. An Informatics Approach to Implement Support for Shared Decision Making for Primary Prevention Statin Therapy. MDM Policy Pract 2018;3(1):2381468318777752 [FREE Full text] [doi: 10.1177/2381468318777752] [Medline: 30288449]
21. Ballard AY, Kessler M, Scheitel M, Montori VM, Chaudhry R. Exploring differences in the use of the statin choice decision aid and diabetes medication choice decision aid in primary care. BMC Med Inform Decis Mak 2017 Aug 10;17(1):118 [FREE Full text] [doi: 10.1186/s12911-017-0514-5] [Medline: 28797295]
22. Percefull J, Butler J. Improving mammography through effective screening, brief intervention, and referral to treatment at a rural health center. J Am Assoc Nurse Pract 2020 Jan 16:324-330. [doi: 10.1097/JXX.0000000000000356] [Medline: 31972786]
23. Denig P, Schuling J, Haaijer-Ruskamp F, Voorham J. Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial. BMJ 2014 Sep 25;349:g5651 [FREE Full text] [doi: 10.1136/bmj.g5651] [Medline: 25255799]

XSL•FO
RenderX

24. Schott SL, Xu K, Berkowitz J, Petersen C, Saunders C, Sobti N, et al. Timing of electronic health record integrated decision aid (IDEA) for stroke prevention in atrial fibrillation matters. Journal of the American College of Cardiology 2019 Mar;73(9):3005. [doi: 10.1016/s0735-1097(19)33611-3]
25. Ajeesh S, Luis R. A Comprehensive Electronic Health Record Based Patient Navigation Module Including Technology Driven Colorectal Cancer Outreach and Education. J Cancer Educ 2018 Jun;33(3):627-633. [doi: 10.1007/s13187-017-1184-9] [Medline: 28188568]
26. Wan Q, Makeham M, Zwar NA, Petche S. Qualitative evaluation of a diabetes electronic decision support tool: views of users. BMC Med Inform Decis Mak 2012 Jul 03;12:61 [FREE Full text] [doi: 10.1186/1472-6947-12-61] [Medline: 22759239]
27. Finkelstein J, Wood J, Crew K, Kukafka R. Introducing a comprehensive informatics framework to promote breast Cancer risk assessment and chemoprevention in the primary care setting. AMIA Summits on Translational Science Proceedings Jul 2017 Jul 26:58-67. [doi: 10.1109/ichi.2016.41] [Medline: 28815107]
28. Richardson S, Feldstein D, McGinn T, Park LS, Khan S, Hess R, et al. Live Usability Testing of Two Complex Clinical Decision Support Tools: Observational Study. JMIR Hum Factors 2019 May 15;6(2):e12471 [FREE Full text] [doi: 10.2196/12471] [Medline: 30985283]
29. Jenssen BP, Shelov ED, Bonafide CP, Bernstein SL, Fiks AG, Bryant-Stephens T. Clinical Decision Support Tool for Parental Tobacco Treatment in Hospitalized Children. Appl Clin Inform 2016;7(2):399-411 [FREE Full text] [doi: 10.4338/aci-2015-12-ra-0169] [Medline: 27437049]
30. Kuo AM, Thavalathil B, Elwyn G, Nemeth Z, Dang S. The Promise of Electronic Health Records to Promote Shared Decision Making: A Narrative Review and a Look Ahead. Med Decis Making 2018 Nov;38(8):1040-1045. [doi: 10.1177/0272989X18796223] [Medline: 30226100]
31. Scalia P, Durand M, Forcino RC, Schubbe D, Barr PJ, O'Brien N, et al. Implementation of the uterine fibroids Option Grid patient decision aids across five organizational settings: a randomized stepped-wedge study protocol. Implement Sci 2019 Sep 02;14(1):88 [FREE Full text] [doi: 10.1186/s13012-019-0933-z] [Medline: 31477140]
32. Yin R. Case Study Research and Applications Design and Methods. Los Angeles: Sage publications; 2017.
33. Rodgers M, Thomas S, Harden M, Parker G, Street A, Eastwood A. Developing a methodological framework for organisational case studies: a rapid review and consensus development process. In: Health Serv Deliv Res. Southampton: NIHR Journals Library; Jan 2016:1-142 URL: https://www.ncbi.nlm.nih.gov/books/NBK333581/pdf/Bookshelf_NBK333581.pdf
34. DynaMed Shared Decisions. EBSCO Health. URL: https://health.ebsco.com/products/option-grid [accessed 2020-06-01]
35. What is SMART? SMART Health IT. URL: https://smarthealthit.org/an-app-platform-for-healthcare/about/ [accessed 2020-06-01]
36. HL7. URL: https://www.hl7.org/ [accessed 2020-06-01]
37. Kamel PI, Nagy PG. Patient-Centered Radiology with FHIR: an Introduction to the Use of FHIR to Offer Radiology a Clinically Integrated Platform. J Digit Imaging 2018 Jun;31(3):327-333 [FREE Full text] [doi: 10.1007/s10278-018-0087-6] [Medline: 29725963]
38. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep:899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
39. Semenov I, Kopanitsa G, Denisov D, Alexandr Y, Osenev R, Andreychuk Y. Patients Decision Aid System Based on FHIR Profiles. J Med Syst 2018 Jul 31;42(9):166. [doi: 10.1007/s10916-018-1016-4] [Medline: 30066031]
40. Melnick ER, Lopez K, Hess EP, Abujarad F, Brandt CA, Shiffman RN, et al. Back to the Bedside: Developing a Bedside Aid for Concussion and Brain Injury Decisions in the Emergency Department. EGEMS (Wash DC) 2015;3(2):1136. [doi: 10.13063/2327-9214.1136] [Medline: 26290885]
41. Jetelina KK, Woodson TT, Gunn R, Muller B, Clark KD, DeVoe JE, et al. Evaluation of an Electronic Health Record (EHR) Tool for Integrated Behavioral Health in Primary Care. J Am Board Fam Med 2018;31(5):712-723 [FREE Full text] [doi: 10.3122/jabfm.2018.05.180041] [Medline: 30201667]
42. McCrorie C, Benn J, Johnson OA, Scantlebury A. Staff expectations for the implementation of an electronic health record system: a qualitative study using normalisation process theory. BMC Med Inform Decis Mak 2019 Nov 14;19(1):222 [FREE Full text] [doi: 10.1186/s12911-019-0952-3] [Medline: 31727063]
43. Scalia P, Elwyn G, Durand M. "Provoking conversations": case studies of organizations where Option Grid™ decision aids have become 'normalized'. BMC Med Inform Decis Mak 2017 Aug 18;17(1):124 [FREE Full text] [doi: 10.1186/s12911-017-0517-2] [Medline: 28821256]
44. Silvia KA, Sepucha KR. Decision aids in routine practice: lessons from the breast cancer initiative. Health Expect 2006 Sep;9(3):255-264. [doi: 10.1111/j.1369-7625.2006.00393.x] [Medline: 16911140]
45. Uy V, May SG, Tietbohl C, Frosch DL. Barriers and facilitators to routine distribution of patient decision support interventions: a preliminary study in community-based primary care settings. Health Expect 2014 Jul;17(3):353-364 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00760.x] [Medline: 22212453]

XSL•FO
RenderX

46. Bloomfield RA, Polo-Wood F, Mandel JC, Mandl KD. Opening the Duke electronic health record to apps: Implementing SMART on FHIR. Int J Med Inform 2017 Dec;99:1-10. [doi: 10.1016/j.ijmedinf.2016.12.005] [Medline: 28118917]

Abbreviations

AI: artificial intelligence
EBSCO: Elton B Stephens Company
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
LLC: limited liability company
PCORI: Patient-Centered Outcomes Research Institute
PDA: patient decision aid
PHI: protected health information
SMART: Substitutable Medical Applications and Reusable Technologies
UPFRONT: Uterine Fibroids Options

Edited by R Kukafka; submitted 22.07.20; peer-reviewed by L Lenert, A Hoffman; comments to author 01.09.20; revised version received 20.10.20; accepted 17.02.21; published 03.05.21.

Please cite as:
Scalia P, Ahmad F, Schubbe D, Forcino R, Durand MA, Barr PJ, Elwyn G
Integrating Option Grid Patient Decision Aids in the Epic Electronic Health Record: Case Study at 5 Health Systems
J Med Internet Res 2021;23(5):e22766
URL: https://www.jmir.org/2021/5/e22766
doi:10.2196/22766
PMID:33938806

©Peter Scalia, Farhan Ahmad, Danielle Schubbe, Rachel Forcino, Marie-Anne Durand, Paul James Barr, Glyn Elwyn. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2021. 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

Patient Portal Use, Perceptions of Electronic Health Record Value, and Self-Rated Primary Care Quality Among Older Adults: Cross-sectional Survey

Dori A Cross[1], PhD; Zachary Levin[1], BA; Minakshi Raj[2], MPH, PhD

[1]Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN, United States

[2]Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, United States

Corresponding Author:
Dori A Cross, PhD
Division of Health Policy and Management
University of Minnesota School of Public Health
420 Delaware Ave SE
MMC 729
Minneapolis, MN, 55455
United States
Phone: 1 612 301 2570
Email: dcross@umn.edu

Abstract

Background: Older adults are increasingly accessing information and communicating using patient-facing portals available through their providers' electronic health record (EHR). Most theories of technology acceptance and use suggest that patients' overall satisfaction with care should be independent of their chosen level of portal engagement. However, achieving expected benefits of portal use depends on demonstrated support from providers to meet these expectations. This is especially true among older adults, who may require more guidance. However, little is known about whether misalignment of expectations around technology-facilitated care is associated with lower perceptions of care quality.

Objective: The aims of this study were to analyze whether older adults' assessment of primary care quality differs across levels of patient portal engagement and whether perceptions of how well their provider uses the EHR to support care moderates this relationship.

Methods: We conducted a cross-sectional survey analysis of 158 older adults over the age of 65 (average age 71.4 years) across Michigan using a 13-measure composite of self-assessed health care quality. Portal use was categorized as none, moderate (use of 1-3 functionalities), or extensive (use of 4-7 functionalities). EHR value perception was measured by asking respondents how they felt their doctor's EHR use improved the patient–provider relationship.

Results: Moderate portal users, compared to those who were extensive users, had lower estimated care quality (–0.214 on 4-point scale; *P*=.03). Differences between extensive portal users and nonportal users were not significant. Quality perception was only particularly low among moderate portal users with low EHR value perception; those with high EHR value perception rated quality similarly to other portal user groups.

Conclusions: Older adults who are moderate portal users are the least satisfied with their care, and the most sensitive to perceptions of how well their provider uses the EHR to support the relationship. Encouraging portal use without compromising perceptions of quality requires thinking beyond patient-focused education. Achieving value from use of patient-facing technologies with older adults is contingent upon matched organizational investments that support technology-enabled care delivery. Providers and staff need policies and practices that demonstrate technology adeptness. Older adults may need more tailored signaling and accommodation for technology to be maximally impactful.

(J Med Internet Res 2021;23(5):e22549) doi:10.2196/22549

KEYWORDS

patient portals; electronic health records; primary care; older adults; patient satisfaction

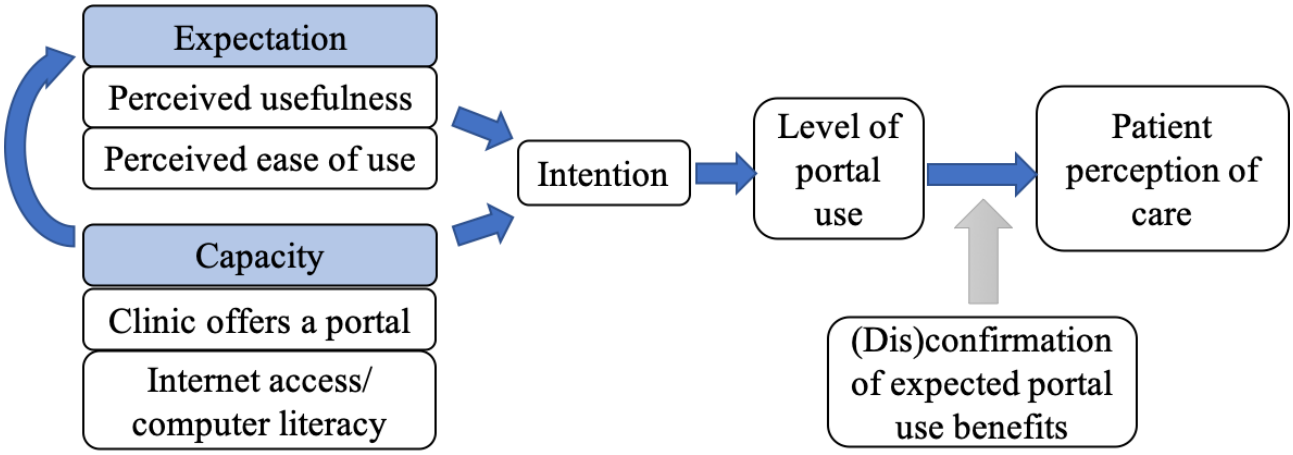
XSL•FO
RenderX

Introduction

Patients' increased access to, and engagement with, their digitized health information has featured prominently in recent federal law and quality-based payment programs designed to accelerate value-generating use of health information technology [1,2]. The Medicare and Medicaid Electronic Health Record (HER) Incentive Programs have helped to advance use of patient portal technologies as a mechanism for information access and electronic patient–provider communication. Portals are intended to increase patient engagement in managing their own health and health care needs, and improve accessibility and efficiency of receiving certain clinical advice, reminders, and services [3-5]. However, evidence linking portal use to improved measures of quality or value is mixed [6-9]. This uncertainty regarding the value of portal use is especially true for older adults, a population for which there has been substantial attention concerning lag in adoption [10] and less understanding of variation in patterns of portal use and perceptions of information technology–enabled care [11-14]. More recent evidence suggests some narrowing of the digital divide based on age [12,15] and significant potential for enhanced communication and self-management engaging older adults through portal technologies [13,16,17]. However, the "right" level of portal use to encourage among older adults is difficult to pinpoint and context dependent.

Patients, provided that they are offered portal access in the first place [15], make usage choices based on technology expectations (eg, perceived usefulness and ease of use) and enabling factors, such as the level of support, encouragement, and education received from providers on how to engage with the portal [18-20]. Together, enabling factors and expectations shape intention and actual portal use, as articulated by the technology acceptance model and the more updated Unified Theory of Acceptance and Use of Technology [21,22] (see Figure 1). However, the linkage between level of portal use and patients' overall satisfaction with care is ill defined.

Figure 1. Integrated conceptual model of portal use and perception of care quality modified by expectation confirmation.



Empirically, there appears to be little association between how much a portal is used (ie, a dose response) and objective measures of quality (eg, hemoglobin A_{1c} control, readmissions) [23,24]. A key theoretical consideration is that portals are a shared tool between patients and providers. A patient's perception of benefit from using the portal is reevaluated with every experience the user has in using the portal to facilitate communication and needed services with a health care provider. Whether a patient is happy with care at their chosen level of portal engagement is therefore dependent on whether the experience matches their expectation [25,26]. Extending the technology acceptance model with expectation–confirmation theory (see Figure 1) reflects this more interactive understanding [27]. (Dis)confirmation of patient expectations around the use of portals, shaped by how the provider organization is engaging from their end, may modify the relationship between chosen levels of engagement and patients' self-assessment of care quality.

Among older adults, the extent to which their level of portal engagement is matched by perceived reciprocal investment by their provider is a heightened contextual factor that may influence how older adults assess their quality of care. Older adults' expectations for how their primary care providers can structure and deliver health care in a more technology-advanced way continue to evolve, and their perceptions of this alignment in investment is heightened given the additional guidance and support that older adults more often need in order to feel comfortable in engaging with patient-facing technologies [28,29]. Observations about how the provider uses the EHR in-clinic—perceived adeptness of use, continuity or disruption of attention and communication, and whether providers voice frustration with use—meaningfully shape whether patients perceive that their provider's use of technology adds value to, or detracts from, the value of their care [30-32]. To the extent that older adults' value assessments around portal engagement may be tied to perceptions of how effectively their providers use the EHR, we must consider these as possibly interdependent factors that shape overall assessment of care quality [29,31,33].

This study used a novel dataset of older adults surveyed about their health needs, provider relationships, and perceptions of health system experiences. We characterized the nature of portal engagement among a sample of older adults and then addressed the following 2 specific research questions: (1) Are older adults' assessments of primary care quality independent of the level of

XSL•FO
RenderX

personal engagement with patient portal technologies? (2) To what extent do patient perceptions of how their provider uses technology modify observed relationships between extent of patient portal use and assessment of care quality? Ultimately, this paper seeks to provide actionable insights that help providers leverage health information technology investment for improved experiences of primary care for a large and growing older adult population.

Methods

Setting and Data Source

A statewide survey regarding older adults' perceptions of health care needs and navigation of primary care services was developed and administered between March 2019 and June 2019 across Michigan, which has population demographics that closely represent nationwide aging trends [34]. The survey instrument was informed by 4 focus groups with a total of 18 older adult participants, and drew upon elements of previously validated surveys, including the clinician/group survey of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) and the National Poll on Healthy Aging [35,36]. "Think aloud" cognitive testing was then conducted with an additional 10 participants for survey refinement to ensure functionality, appropriate wording, and comprehension [37,38]. We used this adaptive measure development approach because many validated measures of primary care quality are general to the adult population; this broader survey effort sought to define quality from the specific perspective of aging. [39] This survey study was reviewed and determined to be exempt from full human subjects review by the University of Michigan Institutional Review Board. The full survey instrument is available upon request.

Participants were recruited statewide using an online health research platform managed by a large academic medical center with a pool of roughly 50,000 prospective participants [40]. Respondents needed to be over 65 years of age, comfortable with written English, have a person whom they consider their "regular doctor", and have passed a brief cognitive test. All respondents must have received health care in the United States at least once in the past year but not necessarily with their regular doctor. Because we used a convenience sampling approach, we collected a number of demographic and health status indicators to help characterize our final sample and appropriately limit the generalizability of our findings [14,16]. Participants received a brief overview of the study and the URL for accessing the survey via Qualtrics where they also provided written informed consent. Respondents were asked to reflect specifically on health care experiences with the person they consider their primary care provider. Survey sections containing the questions of interest for our analyses were distanced from each other within the instrument, and question order within each section was randomized across participants [41]. Participants received US $10 for participation.

Measures

Outcomes

Our primary outcome was a composite measure of participants' overall self-rated health care quality. The survey contained 13 questions that assessed dimensions of quality. Four of these questions were adopted from the CAHPS patient survey [36]; 9 additional questions assessing care coordination, patient-centered care experience, and age-sensitive care delivery were added based on focus group findings to reflect the particular experiences of older adults (Multimedia Appendix 1). All questions were assessed on the same 1-4 scale ("not true," "somewhat true," "fairly true," or "very true") as preferred by respondents during cognitive testing and reverse coded as necessary. We averaged the values provided across all 13 questions to assign each respondent a summary value between 1 and 4. The Cronbach α reliability coefficient was .94, and exploratory factor analysis supported use of a single factor, with all questions loading on to this factor at or above 0.68 (Multimedia Appendix 2).

Technology Perception and Use

We used 2 health information technology–related survey questions. One measure captured how respondents use health information technology via available patient portal features. We asked about usage of 7 features across 3 domains: information availability (viewing laboratory/test results, reviewing physician's advice, finding medications), communication (messaging with provider, answering previsit questions), and convenience (refilling medications, paying bills online). These questions were adapted from the National Poll on Healthy Aging [35]. We then created 3 categories of portal use: no portal use, moderate portal use (1-3 features used), and extensive portal use (4-7 features used).

The second question focused on perception of provider's use of EHRs. We asked respondents, on a 1-4 scale, if they felt their doctor's EHR use improved their patient–provider relationship. We focused on improved relationship as a measure of EHR use given our interest in the value of EHR for communication and patient engagement. We created a binary indicator for EHR value perception; participants who reported a 1 or 2 ("not true" or "somewhat true") were labeled as having a low EHR value perception and those who reported a 3 or 4 ("fairly true" or "very true") were labeled as having a high EHR value perception.

Covariates

The following self-reported demographic and clinical profile characteristics were available from survey data: age, race, sex, highest education completed, and an indicator of financial status (how often respondent has money left over at the end of the month). Clinical profile measures included the patient's self-reported health status (poor or fair, good, and very good or excellent), an indicator of polypharmacy (ie, whether a participant self-reports taking 4 or more pills) [42], number of primary care and number of total health care encounters in the past 6 months, and whether the participant has a caregiver. We also asked respondents the duration of their relationship with their doctor (<5 years, 5-10 year, 10+ years).

XSL•FO
RenderX

Check for Common Method Bias

Because our main variables of interest (ie, information technology perception/use questions) and our composite quality outcome measure derived from the same survey instrument, we tested for common method bias prior to conducting any analyses [41]. Factor analysis on the combination of all included measures confirmed that there was no single latent factor accounting for most of the covariance among survey questions.

Analytic Methods

We first calculated summary statistics for all available demographic and clinical profile characteristics of our survey respondents, and, where possible, considered comparative national statistics to understand how well our sample compared to the demographic profile of older adults across the United States. We also calculated summary statistics for our outcome measure: the 13-item composite of self-reported health care quality.

For our information technology–specific questions, we calculated the percent of respondents using each of the 7 available portal features. To understand the distinguishing characteristics of respondents at each level of portal use (eg, nonportal users, moderate portal users, extensive portal users), we assessed bivariate relationships between portal use and all available demographic and health status measures (listed above as covariates). We also tested the association between our portal use categories and the EHR value perception measure (ie, how respondents perceive value in their provider's use of the EHR). Because health status and health care needs may influence the specific portal features considered useful to an individual, we also ran a supplemental analysis to look at whether the types of portal use (ie, which features) varied based on self-rated health status and/or extent of medication use.

We next examined the bivariate relationships between all available respondent characteristics and the composite quality measure to assess which characteristics are relevant to include in our full models. We estimated full multivariate ordinary least squares regression models with robust SEs. Examining the distribution of errors postestimation supported this model choice. We first estimated the independent adjusted effects of the extent of respondent portal use and the EHR value perception on the composite quality outcome measure. We then reran these models and included an interaction between these variables to assess whether EHR value perception moderates the association between respondents' personal portal use and self-reported health care quality.

Results

Summary of Data

Our survey yielded 167 responses from adults over the age of 65, with 158 surveys sufficiently complete for full analysis. The average age of respondents in our sample was 71.4 years. Respondents were 66.5% (105/158) female, highly educated (45.2% [71/157] reporting higher than a bachelor's degree), and in good health (57.6% [91/158] reporting very good or excellent health status; Table 1). These characteristics, in addition to a younger-skewed age distribution and underrepresentation of rural respondents, differentiate our sample relative to national corresponding statistics. About one-third of participants (46/153, 30.1%) had a formal or informal caregiver who supported their health and health care needs, and 17.9% (28/158) expressed concern with technology use. Individuals varied in clinical complexity (54/157 [34.3%] taking 4 or more pills daily) and financial security (53/158 [33.5%] reporting money rarely or sometimes left over at the end of the month). Over half (88/158, 55.7%) of the respondents had been seeing their doctor for at least 5 years.

Our 13-item composite outcome measure of perceived health care quality averaged 3.46 out of 4 across our sample (SD 0.46). Of the 158 respondents, 30 (19.0%) reported uniformly high quality on all 13 questions.

Portal use among this sample was relatively high. Portal use was found to be positively associated with the number of doctors a patient had seen in a 6-month time frame and negatively associated with age, technology concerns, and financial stability (Multimedia Appendix 3). Most health status characteristics (eg, self-reported health status, polypharmacy, having a caregiver) were not associated with extent of portal use. Of all respondents (N=164 with portal use data), 127 (77.4%) used at least 1 of 7 available functionalities (Figure 2). The most commonly used features were viewing test/laboratory results (72.6%, 119/164) and communicating with the doctor (56.1%, 92/164). The average respondent used 3 different functions; 22.6% (37/164) of the sample reported no portal use, 30.4% (50/164) used 1-3 features (moderate), and 47.0% (77/164) of the sample used 4 or more features (extensive). We found that individuals taking more medications were more likely to use the portal to find their medication list but less likely to use the portal for answering previsit questions. We observed no significant differences by self-reported health status (Multimedia Appendix 4).

Overall, 67 respondents (41.5%) reported that the EHR improved their relationship with their physician (high EHR value perception; Figure 3). This percentage was greatest among extensive portal users (42/76, 55%). Respondents who were nonusers and moderate portal users were significantly less likely to have high EHR value perception (nonusers: 9/34, 26%; moderate users: 16/49, 33%; chi-square *P* value=.01).

XSL•FO
RenderX

Table 1. Respondents' sample characteristics.

Characteristic	Survey sample (N=158), n (%)	Corresponding national statistics (%)
Demographics		
Age (years)[a]		
65-69	83 (52.5%)	32.6%
70-74	35 (22.2%)	25.6%
75-79	24 (15.2%)	17.7%
80+	16 (10.1%)	24.2%
Female[a]	105 (66.5%)	55.5%
Nonmetropolitan county[b]	9 (5.7%)	14.1%
Highest education attained[c]		
< Bachelor's degree	41 (26.1%)	70.8%
Bachelor's degree	45 (28.7%)	16.5%
> Bachelor's degree	71 (45.2%)	12.7%
Has money left over at the end of the month (always/often)[d]	105 (66.5%)	66.7%
Concerns with technology use	28 (17.9%)	—[e]
Health status		
Self-reported health status[f]		
Fair/poor	29 (18.4%)	23.0%
Good	38 (24.1%)	33.1%
Very good/excellent	91 (57.6%)	43.9%
Percent of patients taking 4+ pills daily[g] (n=157)	54 (34.3%)	54%
Has a caregiver (n=153)	46 (30.1%)	—
Patient–provider relationship		
Seeing personal doctor for 5+ years	88 (55.7%)	—

[a]National statistics from the US Census Bureau, Population Division (2019) [43].

[b]National statistics from the US Department of Agriculture Economic Research Service (2019) [44].

[c]National statistics from the US Census Bureau (2018) [45].

[d]National statistics from the National Council on Aging (2015) [46].

[e]Comparable national statistics could not be found.

[f]National statistics from the Centers for Disease Control and Prevention (2018) [47].

[g]National statistics from the Kaiser Family Foundation (2019) [48].

XSL•FO
RenderX

Figure 2. Use of available portal functions among respondents.

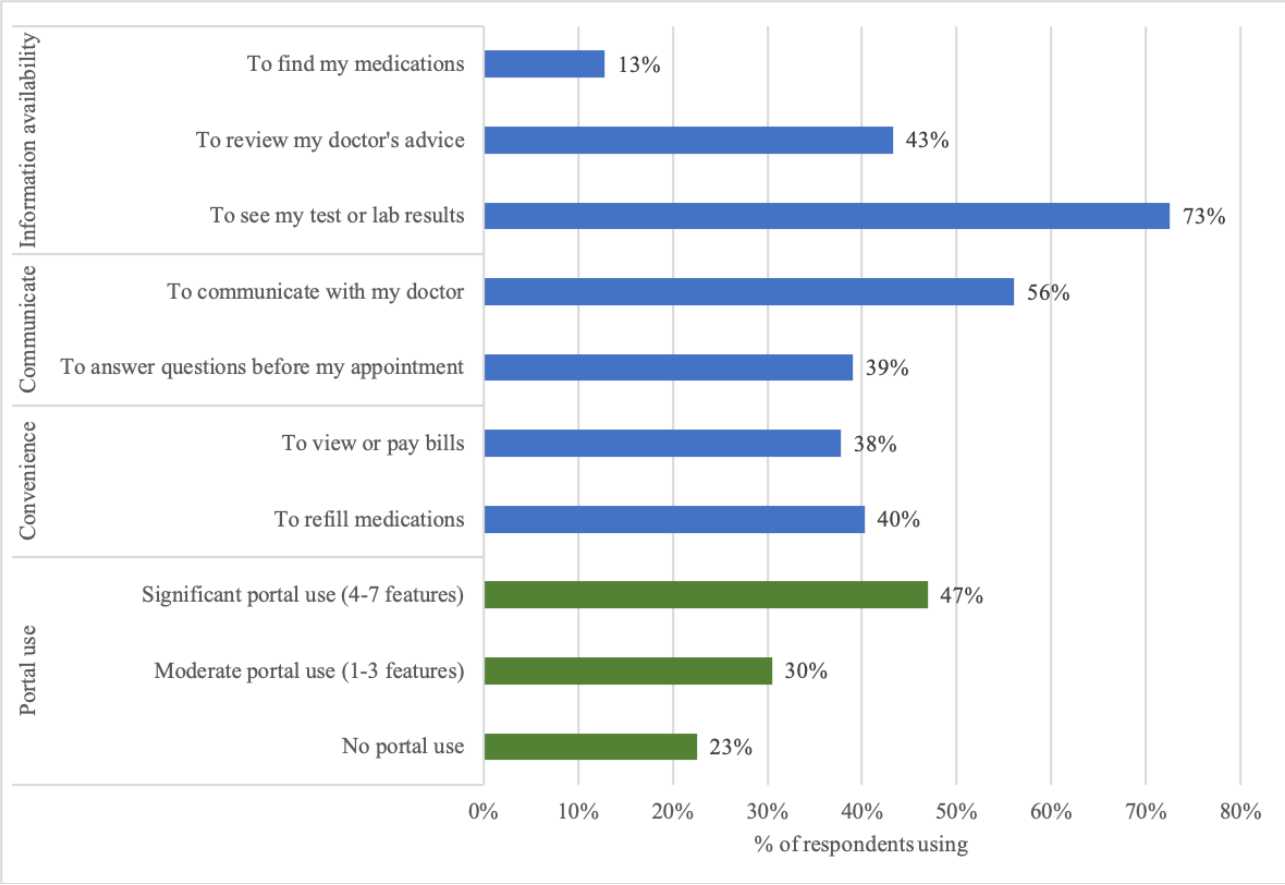
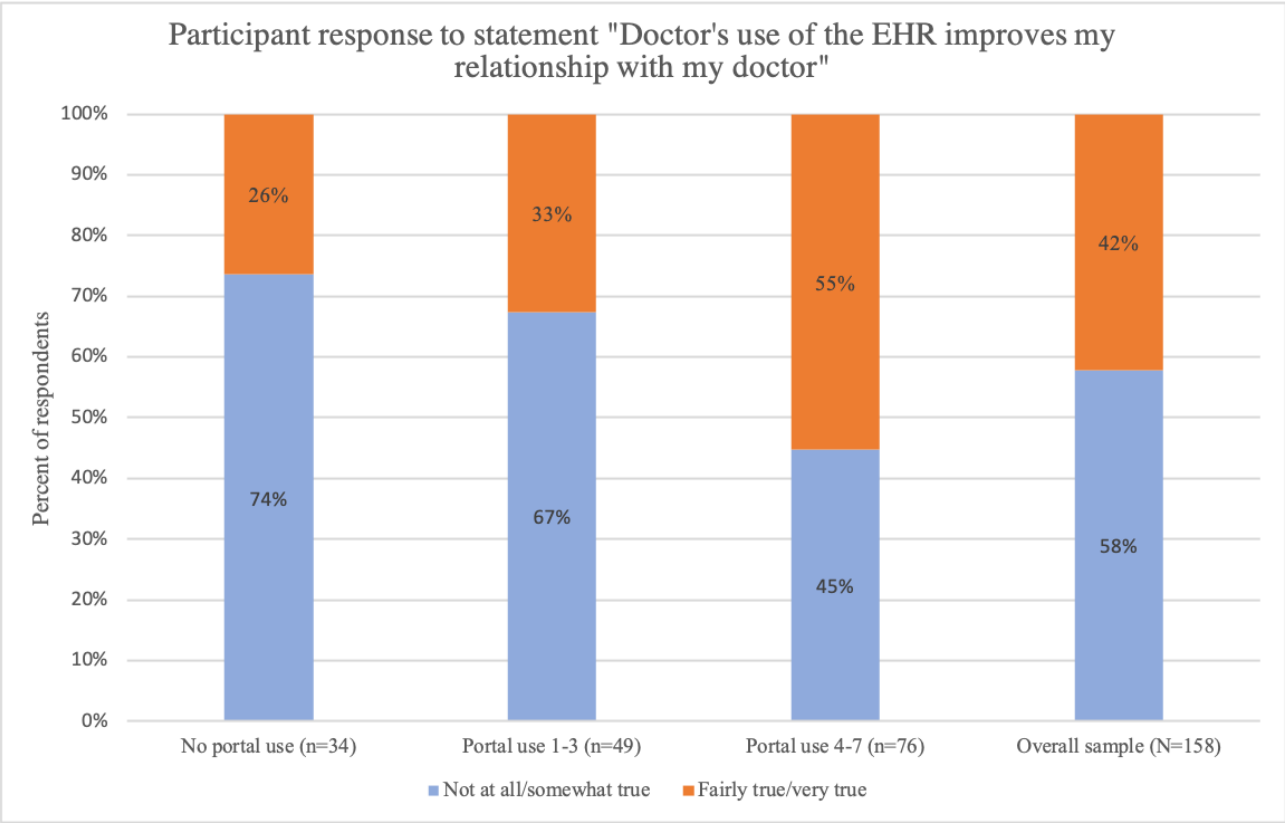


Figure 3. Comparison of EHR value perception by portal use category. EHR: electronic health record.



XSL•FO
RenderX

Regression Results

Bivariate analyses that tested association of each patient-level covariate and the composite quality outcome measure revealed only 3 significant associations: respondents' length of relationship with their provider, self-reported health, and whether the respondent had money leftover at the end of the month. The multivariate models presented include only these controls. Models with all available covariates included had consistent findings; because of our sample size, we present the more parsimonious model.

Column 1 of Table 2 presents our noninteracted multivariate model. Relative to extensive portal users, moderate portal use was significantly associated with lower self-rated quality (–0.214 on 4-point composite scale; *P*=.03). Adjusted differences between extensive portal users and nonportal users were not significant. Estimates of self-reported quality were also significantly higher for individuals with high EHR value perception relative to low perception (0.288; *P*<.001). Among control variables, higher quality was associated with individuals having a longer (10+ years) relationship duration with their provider and with being in very good or excellent health. Effects of financial stability were not significant.

When we interacted EHR value perception with respondent portal use (column 2 of Table 2), we still observed high EHR value perception to be positively and significantly associated with self-rated quality, and that there were significant differences in quality only between extensive and moderate portal users (not between extensive and nonportal users). The interaction terms were not statistically significant, suggesting that these significant differences in self-rated quality by portal use category are more salient among those with low EHR value perception. Figure 4 visually demonstrates the effect of EHR value perception on the relationship between portal use and self-rated quality. In the noninteracted model (ie "independent effects"), the effect of EHR value perception remained consistent across all levels of portal use. The interacted model demonstrates differential effects where moderate portal users appear particularly sensitive to EHR value perception. Among this group, those with low EHR value perception had especially low predicted estimates of perceived quality, while those with high EHR value perception reported a more similar self-rated quality to those who were extensive portal users.

Table 2. Multivariate ordinary least squares regression results: adjusted estimated effects on overall patient self-rated quality (N=158).[a]

Estimated effects on self-rated quality	Model without interaction, coefficient (SE)	*P* value	Model with interaction, coefficient (SE)	*P* value
Portal use category (reference: extensive portal use)				
No portal use	–0.081 (0.104)	.434	–0.041 (0.121)	.73
Moderate portal use	–0.214 (0.095)	.027	–0.288 (0.131)	.03
High EHR[b] value perception (reference: low EHR value perception)	0.288 (0.074)	<.001	0.257 (0.086)	.004
Portal use x EHR value perception (reference: high EHR value perception x extensive portal use)				
High EHR value perception x no portal use	N/A[c]	N/A	–0.181 (0.228)	.44
High EHR value perception x moderate portal use	N/A	N/A	0.221 (0.181)	.23
Length of relationship with primary care physician (reference: <5 years)				
5-10 years	0.176 (0.097)	.071	0.173 (0.097)	.08
10+ years	0.350 (0.086)	<.001	0.339 (0.086)	<.001
Self-reported health (reference: very good/excellent)				
Poor/fair	–0.246 (0.119)	.041	–0.221 (0.123)	.08
Good	–0.216 (0.092)	.034	–0.196 (0.092)	.06
Money left over at the end of the month always/often (reference: rarely/sometimes)	0.080 (0.099)	.422	0.091 (0.099)	.36
Constant	3.19 (0.163)	<.001	3.40 (0.151)	<.001
R^2	0.274	N/A	0.288	N/A

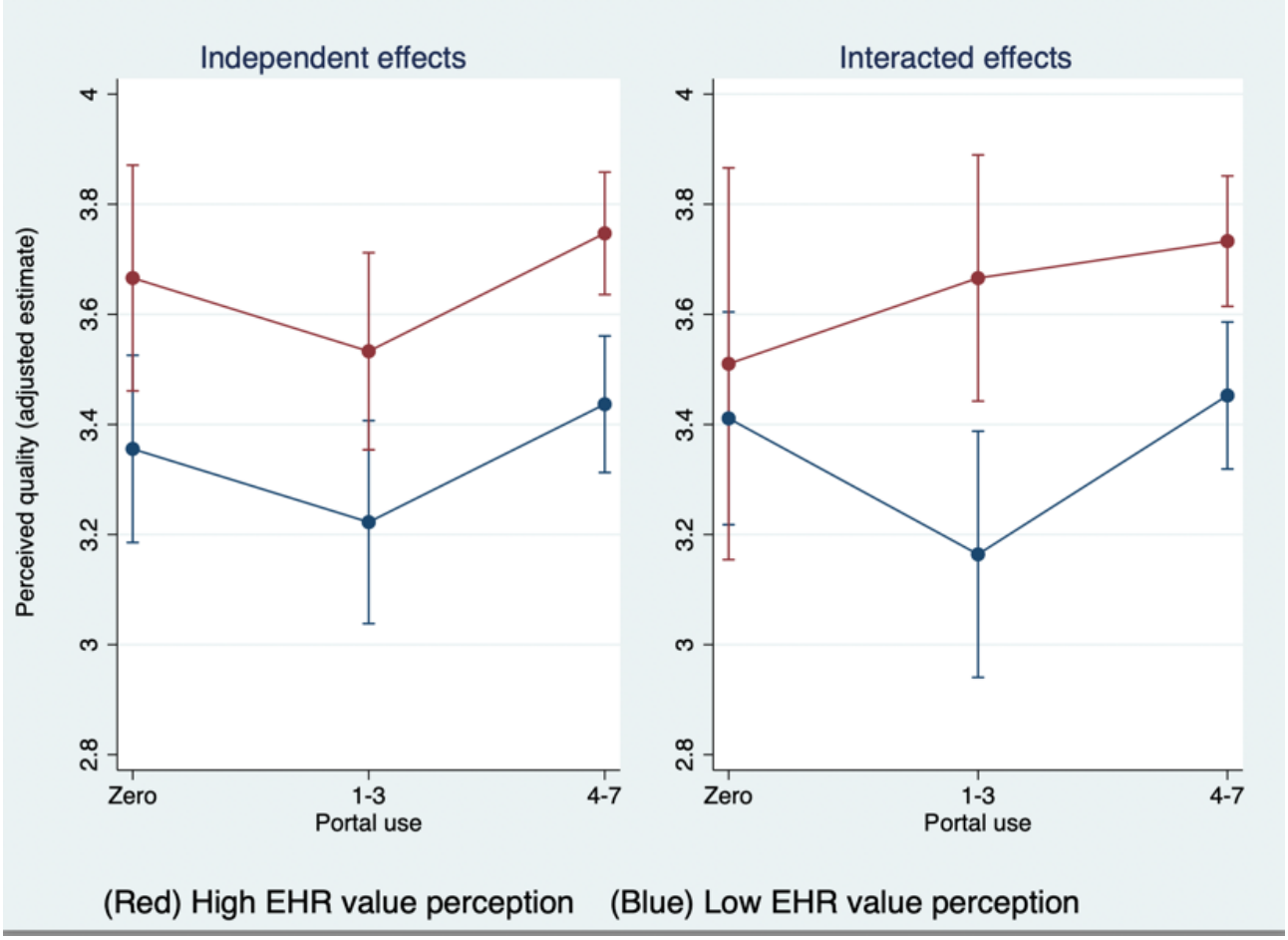
[a]Outcome (self-rated quality) on a 4-point scale.

[b]EHR: electronic health record.

[c]N/A: not applicable.

XSL•FO
RenderX

Figure 4. Adjusted estimates of self-rated quality by EHR value perception and portal use. EHR: electronic health record.



Discussion

Principal Findings

Using novel statewide survey data of older adults' health care needs and preferences, we found that moderate portal users, those who use just a few available features, had the lowest self-rated quality, lower than both extensive portal users and nonportal users. Among this group, it was particularly those individuals also with low EHR value perception who had the lowest estimates of self-rated quality. Those with high EHR value perception had estimated quality levels similar to those of extensive portal users. These analyses preliminarily support an expectancy confirmation theory extension to patient technology acceptance. Our findings are a first step toward understanding how a mismatch between older adults' chosen level of engagement in technology-enabled care and perceptions of how well their provider uses technology to deliver services may drive variation in how older adults assess overall quality of their primary care.

Our results reveal an important emerging dual consideration with older adults. The number of older adults who are interested in and capable of participating in technology-supported health care is growing especially as it enables them to age in place [49,50]. Even those with some trepidation are starting to engage with these tools, and thoughtful support of this experience is increasingly a key component of how they perceive their care experience. These considerations now exist alongside longstanding concerns for a significant number of older adults who still have very salient individual and structural barriers to engagement (to the point that they are not even meaningfully captured via a primarily online survey like what has been presented in this study). Both issues merit attention and investments in improvement.

Fostering higher and more equitable rates of portal engagement with older adults requires offering portal access more equitably [15] and coaching patients (and their family members) through the features while being more explicit about shared expectations and the "rules of engagement" regarding portal use [13,20]. Encouraging patient portal use without compromising perceptions of quality, however, also requires that clinics demonstrate that they are competent, proactive users of technology. This means, first, that providers and staff must convey their own commitment to using the portal as a communication tool, including more robust provider or staff training and explicit policies around responsiveness and follow-up [20,29]. Enabling and sustaining this commitment requires health care organizations to support providers in these efforts, for example, by building time for portal encounters into scheduling and productivity considerations, and using EHR features such as shared task queues to facilitate delegation of communications that do not require physician response [51].

Our results also underscore the importance of opinions that patients form regarding providers' adeptness at using the EHR during in-person care. These opinions—in particular for those

XSL•FO
RenderX

patients who are engaging with the portal but are not quite super users—are strongly associated with perceptions of care quality. Significant EHR usability challenges and documentation burden have persisted as being reasons for why providers are frustrated with using their systems [52,53]. However, patients' unmet expectations may increasingly become a consequence to providers' obvious frustration and disengagement with the EHR. Primary care settings need implementable strategies to convey to patients the ways that technology is being leveraged to support their care, for example, by training and evaluating providers and staff on participative communication strategies that use screen sharing or integration of EHR task completion with patient engagement [31,54]. This may be especially necessary for providers caring for those with mistrust of the health care system and/or those with limited English proficiency or other communication or health literacy barriers, who may benefit from engaging with technology but may be wary of providers' capacity to deliver high value care through the EHR. Policy efforts are also needed to encourage vendors to improve usability and development of features that meaningfully support the most cognitively challenging and time-intensive tasks in primary care [55]. Ongoing federal initiatives that focus on enabling use of third-party applications for self-management and patient-generated health data (ie, tools untethered to the EHR) need to also be accompanied by guidelines and support for integrating these tools in to practice.

Future Work

As older adults' options for technology-assisted health and health care continue to grow, shared expectations are increasingly important for how use of these tools will be supported and matched by adept technology use among providers. Future research should emphasize the salient contextual factors that influence technology use and technology perceptions particularly relevant to older adults' health care utilization. For example, older adults are likely to receive support from family or informal caregivers, and caregivers' perception of provider quality may influence their older relatives' perception of quality. Future research should consider how the informal caregivers' role could be best accommodated and facilitated in the presence of EHR-supported care delivery [56,57]. Simultaneously, providers and organizations should be considerate of the potential for additional burdens on caregivers to foster their older relative's engagement with technology. Organizational changes required to meet these identified needs are not insignificant, but can build the essential capacity for patient-centered, value-based care.

Limitations

This study has two key limitations. First, this survey is a cross-sectional analysis and we cannot make any causal claims about the impact of patient or provider use of technology on older adults' assessment of care quality. Our research approach is descriptive and not designed to address the endogeneity of portal use with respect to health needs and the nature of recent health encounters. Second, the study uses a convenience sample and diverges on some key characteristics (ie, age distribution, educational achievement, rurality) relative to nationwide statistics. The sample also comprised mostly community-dwelling older adults and was primarily accessed online, excluding those older adults living in nursing home facilities, those with severe illnesses, and those with significant technology barriers. Our findings are thus not broadly generalizable, particularly for more clinically and socially complex older adult populations, but do provide important insights for the growing number of older adults who are interested and capable of participating in technology-supported health care.

Conclusions

Despite significant barriers for many (ie, connectivity and individual comfort with technology), certain populations of older adults are using patient portals and updating their expectations about technology-enabled health care delivery at growing rates. This 2019 statewide survey is evidence for widespread but highly variable portal use. Older adults who are moderate portal users are the least satisfied with their care and the most sensitive to perceptions of how well their provider uses the EHR to support care. This offers preliminary evidence to inform an understanding of the link between patient technology use and perceptions of care quality, moderated by expectancy confirmation. Patients' satisfaction with care at different chosen levels of portal use depends on whether providers offer an experience that aligns with their expectations. Encouraging older adults in more nascent use of patient portals may negatively affect perceptions of care quality if providers are not demonstrating adeptness with their own technology use during in-person visits and for asynchronous interaction. Organizationally, clinics need to consider changes to technology-enabled care—such as better access for, and integration of, caregivers—that are sensitive to, and accommodating of, the evolving needs and expectations of older adults. Ultimately, regulatory and payer policy changes are necessary to address the root causes of provider's frustration and disengagement with technology-supported care practices, especially poor EHR usability and lack of support for integration and efficient use of portals or other remote technologies.

Acknowledgments

The authors thank the study respondents for taking the time to participate, and thank Katie Larin for her contributions to this manuscript. The authors also acknowledge the University of Michigan Rackham Graduate School and the Blue Cross Blue Shield Foundation of Michigan for their support of this work.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions for developing the quality composite measure.
[DOCX File , 21 KB - jmir_v23i5e22549_app1.docx]

Multimedia Appendix 2
Exploratory factor analysis.
[DOCX File , 13 KB - jmir_v23i5e22549_app2.docx]

Multimedia Appendix 3
Characteristics associated with portal use.
[DOCX File , 17 KB - jmir_v23i5e22549_app3.docx]

Multimedia Appendix 4
Variation in portal use, by number of medications taken and by self-reported health status.
[DOCX File , 94 KB - jmir_v23i5e22549_app4.docx]

References

1. Lye C, Forman H, Daniel J, Krumholz H. The 21st Century Cures Act and electronic health records one year later: will patients see the benefits? J Am Med Inform Assoc 2018 Sep 01;25(9):1218-1220. [doi: 10.1093/jamia/ocy065] [Medline: 30184156]
2. Centers for Medicare and Medicaid. Quality payment program: promoting interoperability measures requirements. Centers for Medicare and Medicaid. URL: https://qpp.cms.gov/mips/promoting-interoperability [accessed 2021-03-14]
3. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
4. What is a patient portal? Office of the National Coordinator for Health IT. 2020. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2020-06-01]
5. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
6. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
7. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
8. Tenforde M, Jain A, Hickner J. The value of personal health records for chronic disease management: what do we know? Fam Med 2011 May;43(5):351-354 [FREE Full text] [Medline: 21557106]
9. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
10. Peacock S, Reddy A, Leveille S, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 01;24(e1):e173-e177 [FREE Full text] [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
11. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
12. Nahm E, Zhu S, Bellantoni M, Keldsen L, Charters K, Russomanno V, et al. Patient portal use among older adults: what is really happening nationwide? J Appl Gerontol 2020 Apr;39(4):442-450 [FREE Full text] [doi: 10.1177/0733464818776125] [Medline: 29779422]
13. 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]
14. Cajita MI, Nilsen ML, Irizarry T, Callan JA, Beach SR, Swartwout E, et al. Predictors of patient portal use among community-dwelling older adults. Res Gerontol Nurs 2021 Jan 01;14(1):33-42. [doi: 10.3928/19404921-20200918-03] [Medline: 32966584]

XSL•FO
RenderX

15. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
16. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
17. 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]
18. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
19. Irizarry T, DeVito Dabbs A, 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]
20. Sieck CJ, Hefner JL, Schnierle J, Florian H, Agarwal A, Rundell K, et al. The rules of engagement: perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 04;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]
21. Bagozzi RP, Davis FD, Warshaw PR. Development and test of a theory of technological learning and usage. Human Relations 2016 Apr 22;45(7):659-686. [doi: 10.1177/001872679204500702]
22. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
23. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient portals. Appl Clin Inform 2017 Dec 16;07(02):489-501. [doi: 10.4338/aci-2016-01-ra-0003]
24. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
25. Price-Haywood E, Luo Q, Monlezun D. Dose effect of patient-care team communication via secure portal messaging on glucose and blood pressure control. J Am Med Inform Assoc 2018 Jun 01;25(6):702-708 [FREE Full text] [doi: 10.1093/jamia/ocx161] [Medline: 29444256]
26. Parasuraman A, Berry L, Zeithaml V. Understanding customer expectations of service. Sloan Manage Rev 1991;32(3):39-48.
27. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
28. Graetz I, Gordon N, Fung V, Hamity C, Reed M. The digital divide and patient portals. Medical care 2016;54(8):772-779. [doi: 10.1097/mlr.0000000000000560]
29. Kim S, Fadem S. Communication matters: Exploring older adults' current use of patient portals. Int J Med Inform 2018 Dec;120:126-136. [doi: 10.1016/j.ijmedinf.2018.10.004] [Medline: 30409337]
30. Alkureishi MA, Lee WW, Lyons M, Press VG, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
31. Crampton N, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665 [FREE Full text] [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
32. Lelievre S, Schultz K. Does computer use in patient-physician encounters influence patient satisfaction? Can Fam Physician 2010 Jan;56(1):e6-12 [FREE Full text] [Medline: 20090064]
33. Qiao Y, Asan O, Montague E. Factors associated with patient trust in electronic health records used in primary care settings. Health Policy and Technology 2015 Dec;4(4):357-363. [doi: 10.1016/j.hlpt.2015.08.001]
34. QuickFacts: M. US Census Bureau. US Census Bureau. 2017. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2020-06-01]
35. Clark S, Singer D, Solway E, Kirch M, Malani P. Logging in: using patient portals to access health information. National Poll on Healthy Aging, University of Michigan. 2018. URL: https://www.healthyagingpoll.org/report/logging-using-patient-portals-access-health-information [accessed 2021-03-14]
36. CAHPS patient experience surveys and guidance. Agency for Healthcare Research and Quality. 2019. URL: https://www.ahrq.gov/cahps/surveys-guidance/index.html [accessed 2020-06-10]
37. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. Public Opinion Quarterly 2007 Jun 05;71(2):287-311. [doi: 10.1093/poq/nfm006]
38. Daugherty S, Harris-Kojetin L, Squire C, Jaël E, Harris-Kojetin L. Maximizing the quality of cognitive interviewing data: an exploration of three approaches and their informational contributions. 2001 Presented at: Proceedings of the annual meeting of the American Statistical Association; 2001 August 5-9; American Statistical Association.

XSL•FO
RenderX

39. Raj M. Meeting the Geriatric Care Needs of an Aging Population: Older Adult, Family Caregiver, and Provider Perspectives. In: Meeting the Geriatric Care Needs of an Aging Population: Older Adult, Family Caregiver, and Provider Perspectives. Ann Arbor, MI: University of Michigan; 2020.
40. Health Research Recruitment Platform. University of Michigan Health System. Health research. 2020. URL: UMHealthResearch.org [accessed 2019-01-01]
41. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
42. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatr 2017 Oct 10;17(1):230 [FREE Full text] [doi: 10.1186/s12877-017-0621-2] [Medline: 29017448]
43. Annual estimates of the resident population for selected age groups by sex for the United States. US Census Bureau, Population Division. 2019. URL: https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html [accessed 2021-04-01]
44. Pender J, Hertz T T, Cromartie J, Farrigan T. Rural America at a glance, 2019 edition. US Department of Agriculture Economic Research Service. 2019. URL: https://www.ers.usda.gov/publications/pub-details/?pubid=95340 [accessed 2021-04-01]
45. US Census Bureau. Current population survey table creator. URL: https://www.census.gov/cps/data/cpstablecreator.html [accessed 2021-04-01]
46. Get the facts on economic security for seniors. National council on aging. URL: https://www.ncoa.org/article/get-the-facts-on-economic-security-for-seniors [accessed 2021-04-01]
47. Behavioral risk factor surveillance system web-enabled analysts tool. Centers for Disease Control and Prevention. URL: https://nccd.cdc.gov/weat/#/analysis [accessed 2021-04-01]
48. Kirzinger A, Neuman T, Cubanski J, Brodie M. Data note: prescription drugs and older adults. Kaiser Family Foundation. 2019. URL: https://www.kff.org/health-reform/issue-brief/data-note-prescription-drugs-and-older-adults [accessed 2021-04-01]
49. Andrews J, Brown L, Hawley M, Astell A. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. JMIR 2019;21(2):1-11. [doi: 10.2196/11694]
50. Wang J. Mobile and connected health technologies for older adults aging in place. J Gerontol Nurs 2018 Jun 01;44(6):3-5. [doi: 10.3928/00989134-20180509-01] [Medline: 29846735]
51. Cross D. Contextual factors affecting the implementation of team-based primary care: a scoping review. In: Managing Improvement in Healthcare. NY: Palgrave Macmillan; 2018:77-98.
52. Mishra P, Kiang JC, Grant RW. Association of medical scribes in primary care with physician workflow and patient experience. JAMA Intern Med 2018 Nov 01;178(11):1467-1472 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3956] [Medline: 30242380]
53. Committee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. National Academy of Medicine. Washington, D.C: National Academies Press; 2019.
54. Street RL, Liu L, Farber NJ, Chen Y, Calvitti A, Weibel N, et al. Keystrokes, mouse clicks, and gazing at the computer: how physician interaction with the EHR affects patient participation. J Gen Intern Med 2018 Apr;33(4):423-428 [FREE Full text] [doi: 10.1007/s11606-017-4228-2] [Medline: 29188544]
55. Adler-Milstein J. Electronic health record time among outpatient physicians: reflections on the who, what, and why. Ann Intern Med 2020 Jan 14;172(3):212-213. [doi: 10.7326/m19-3921]
56. Wolff JL, Spillman B. Older adults receiving assistance with physician visits and prescribed medications and their family caregivers: prevalence, characteristics, and hours of care. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 1:S65-S72 [FREE Full text] [doi: 10.1093/geronb/gbu119] [Medline: 25342825]
57. Wolff JL, Darer JD, Larsen KL. Family caregivers and consumer health information technology. J Gen Intern Med 2016 Jan;31(1):117-121 [FREE Full text] [doi: 10.1007/s11606-015-3494-0] [Medline: 26311198]

Abbreviations

CAHPS: Consumer Assessment of Healthcare Providers and Systems
EHR: electronic health record

XSL•FO
RenderX

Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by C Johnson, S Bidmon; comments to author 21.09.20; revised version received 09.11.20; accepted 24.03.21; published 10.05.21.

Please cite as:
Cross DA, Levin Z, Raj M
Patient Portal Use, Perceptions of Electronic Health Record Value, and Self-Rated Primary Care Quality Among Older Adults: Cross-sectional Survey
J Med Internet Res 2021;23(5):e22549
URL: https://www.jmir.org/2021/5/e22549
doi:10.2196/22549
PMID:33970111

©Dori A Cross, Zachary Levin, Minakshi Raj. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.05.2021. 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

Evaluating the Effectiveness of NoteAid in a Community Hospital Setting: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Patients

John P Lalor[1], PhD; Wen Hu[2], MSc; Matthew Tran[2]; Hao Wu[3], PhD; Kathleen M Mazor[4,5], EdD; Hong Yu[2,5,6,7], PhD

[1]Department of Information Technology, Analytics, and Operations, Mendoza College of Business, University of Notre Dame, Notre Dame, IN, United States

[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[3]Department of Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, TN, United States

[4]Meyers Primary Care Institute, University of Massachusetts Medical School/Reliant Medical Group/Fallon Health, Worcester, MA, United States

[5]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[6]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States

[7]Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States

Corresponding Author:
Hong Yu, PhD
Department of Computer Science
University of Massachusetts Lowell
Dandeneau Hall
1 University Avenue
Lowell, MA, 01854
United States
Phone: 1 978 934 3620
Email: hong.yu@umassmed.edu

Abstract

Background: Interventions to define medical jargon have been shown to improve electronic health record (EHR) note comprehension among crowdsourced participants on Amazon Mechanical Turk (AMT). However, AMT participants may not be representative of the general population or patients who are most at-risk for low health literacy.

Objective: In this work, we assessed the efficacy of an intervention (NoteAid) for EHR note comprehension among participants in a community hospital setting.

Methods: Participants were recruited from Lowell General Hospital (LGH), a community hospital in Massachusetts, to take the ComprehENotes test, a web-based test of EHR note comprehension. Participants were randomly assigned to control (n=85) or intervention (n=89) groups to take the test without or with NoteAid, respectively. For comparison, we used a sample of 200 participants recruited from AMT to take the ComprehENotes test (100 in the control group and 100 in the intervention group).

Results: A total of 174 participants were recruited from LGH, and 200 participants were recruited from AMT. Participants in both intervention groups (community hospital and AMT) scored significantly higher than participants in the control groups (P<.001). The average score for the community hospital participants was significantly lower than the average score for the AMT participants (P<.001), consistent with the lower education levels in the community hospital sample. Education level had a significant effect on scores for the community hospital participants (P<.001).

Conclusions: Use of NoteAid was associated with significantly improved EHR note comprehension in both community hospital and AMT samples. Our results demonstrate the generalizability of ComprehENotes as a test of EHR note comprehension and the effectiveness of NoteAid for improving EHR note comprehension.

(J Med Internet Res 2021;23(5):e26354) doi:10.2196/26354

KEYWORDS

health literacy; crowdsourcing; natural language processing; information storage and retrieval; psychometrics; intervention; literacy; electronic health record; efficacy; comprehension

XSL•FO
RenderX

Introduction

Access to and demand for heath information has led to a greater focus on patient-centered care [1,2]. Patient-centered care "makes the patient and their loved ones an integral part of the care team who collaborate with health care professionals in making clinical decisions" [2]. While prior work has shown that more active involvement by patients can lead to better outcomes [3,4] and that patients are more proactive than ever in seeking out health information [5,6], challenges remain for patients with low health literacy. The National Assessment of Adult Literacy estimates that approximately 36% of Americans have health literacy levels rated as "basic" or "below basic" [7]. This estimate, combined with the finding that physicians often overestimate the health literacy of their patients [8,9], shows that there is a gap between patient desire for health information and their ability to understand it. This gap can lead to adverse effects for patients as well as higher costs for health care centers [10-16]. Identifying low health literacy individuals and providing resources to improve their understanding are two key areas in population health research [17-20].

One aspect of health literacy that has become more critical in recent years is eHealth literacy [21,22]. As increasing numbers of patients are able to view their medical records online via patient portals (eg, the OpenNotes project) [23], there is a growing need for eHealth literacy tests and interventions to assess and improve eHealth literacy. One tool for assessing eHealth literacy is the ComprehENotes test [24], which tests the ability of individuals to understand free-text notes in a sample of electronic health records (EHRs). The ComprehENotes test consists of multiple-choice questions generated by groups of physicians and medical researchers, and validated using item response theory (IRT). Research has shown that providing access to NoteAid, an educational intervention tool that automatically defines medical terms in lay terms [18,25], improves scores on ComprehENotes items [26]. However, participants in these studies were recruited from the Amazon Mechanical Turk (AMT) platform, and thus may not represent the typical patient population [27,28]. For example, the AMT participants' self-reported demographic characteristics revealed that these participants tended to be younger and better educated than patients at risk of low health literacy [7].

This study examined the impact of NoteAid on participants recruited from Lowell General Hospital (LGH), a community hospital in Massachusetts, including diabetes patients and their friends and family members. We sought to answer the following research question: "Does NoteAid improve EHR note comprehension for participants recruited from a community hospital?" (RQ1).

As a secondary goal, we sought to analyze the differences in performance between participants at the community hospital and participants on the AMT platform. Prior work has shown that NoteAid leads to improved scores on the ComprehENotes test for AMT participants [26]. However, overall scores for all participants on AMT were relatively high, and the population of AMT participants did not include groups typically at higher risk for low health literacy. Therefore, we compared performance on ComprehENotes between participants recruited from a community hospital and participants recruited from AMT to identify differences. The second research question was as follows: "Are participants at a community hospital different from AMT participants in terms of their EHR note comprehension levels as measured by the ComprehENotes test?" (RQ2).

As a third goal, we sought to determine whether NoteAid is equally effective for improving EHR note comprehension between participants at a community hospital and participants on the AMT platform. The third research question was as follows: "Is NoteAid equally effective or differentially effective for community hospital participants as for AMT participants in improving EHR note comprehension?" (RQ3).

Finally, we investigated the performance of different demographic groups on the ComprehENotes test, both with and without NoteAid, to see if ComprehENotes scores vary across subgroups, and whether NoteAid is equally effective across these subgroups. The fourth and fifth research questions were as follows: "Is EHR note comprehension consistent across different demographic groups?" (RQ4) and "Is NoteAid equally effective or differentially effective across different demographic groups?" (RQ5).

Methods

Overview

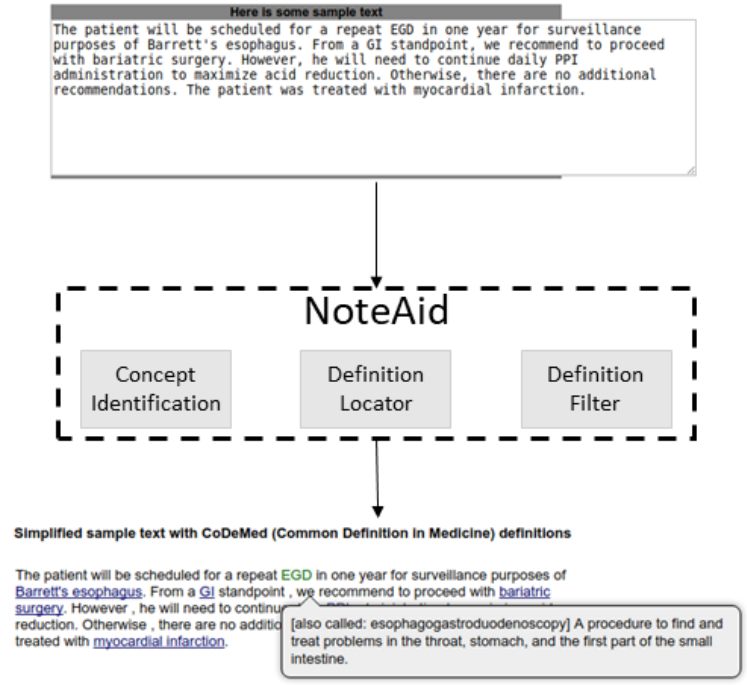
The work in this study was approved by the Institutional Review Boards (IRBs) at the University of Massachusetts Medical School and LGH. All participants from LGH were shown an information sheet describing the study, had the ability to ask questions before participating, and provided verbal informed consent before participating. AMT participants provided electronic informed consent before participating.

NoteAid

NoteAid is a web-based natural language processing (NLP) system for linking medical jargon to lay-language definitions [18,25]. The following two components are central to NoteAid: a repository of lay definitions for medical terms (CoDeMed) and the NLP system for linking medical concepts to these definitions (MedLink). NoteAid is implemented as a web application, where users can navigate to the NoteAid website and enter a snippet of text from their own EHR note. NoteAid will then process the note text and display the note with terms defined via tooltip text. Defined terms are underlined, and users can display a term definition by moving the mouse cursor over the term (Figure 1).

XSL•FO
RenderX

Figure 1. An example of medical terminology definition using NoteAid (image source: [26]).



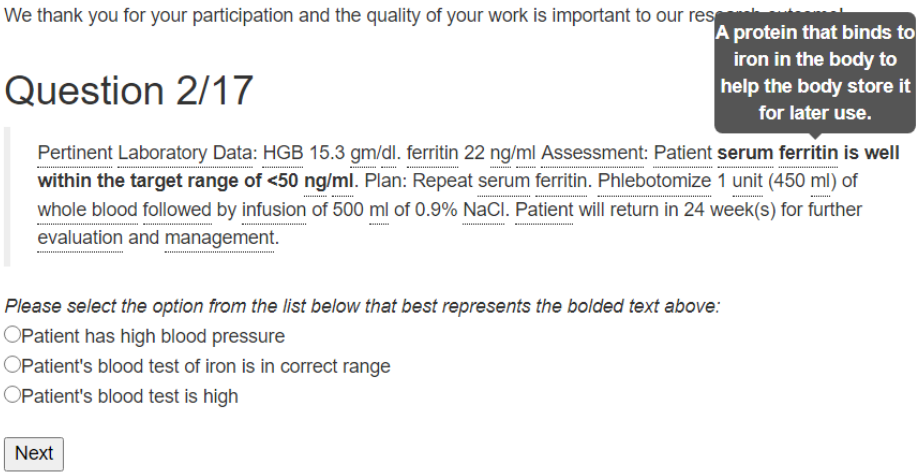
ComprehENotes

Overview

The ComprehENotes test is the first test to directly assess the ability of individuals to comprehend EHR notes. The ComprehENotes test consists of 14 passages taken from deidentified EHR notes. A section of the passage is presented in boldface, and the test takers are asked to select which of three options is the closest in meaning to the bold text (Figure 2). As detailed by Lalor et al [24], the ComprehENotes test was built by a group of physicians and nonclinical medical researchers using the sentence verification technique [29]. Questions were administered to a sample of 660 English-speaking adults on AMT. The psychometric properties of the questions were subsequently analyzed using the IRT method to confirm the performance of the test questions [24,30].

Prior work [24] has shown that ComprehENotes test scores are consistent with demographic expectations with regard to health literacy (ie, less educated respondents score lower than more educated respondents). In addition, providing access to lay definitions for medical terms via NoteAid is associated with higher scores on ComprehENotes [26].

Figure 2. An example of a ComprehENotes test question with embedded NoteAid definitions as implemented on the web application. In this example, the definition of "ferritin" (gray box) is useful in understanding that the bold text describes a blood iron test.



XSL•FO
RenderX

ComprehENotes Administration at a Community Hospital

For the community hospital participants, we implemented the ComprehENotes test as a web application on a tablet, which allowed for flexibility in terms of delivery and intervention modifications. The hardware used was one Microsoft Surface Pro and one Apple iPad Pro 11. When a participant loaded the application, he or she was randomly assigned to either the control or intervention group. Participants provided demographic information on the app, and then proceeded to take the test. The test was administered one question at a time. The order of the questions was randomized. Responses were directly collected and stored on our server. Other than self-reported demographic information, no user information was stored on the server.

ComprehENotes Administration on AMT

For AMT participants, the ComprehENotes test was implemented as a web application where quality control questions were additionally included to ensure that participants completed the task to the best of their ability. Specifically, the AMT participants were given three quality control "check questions" interspersed throughout the ComprehENotes test. These questions were included to ensure that the AMT participants were paying attention as they were completing the task [31]. If participants answered a check question incorrectly, they were presented with a message indicating that they have answered a check question incorrectly, and were given the option to start the task over or exit the window without completing the task.

Integrating NoteAid With ComprehENotes

For the intervention groups (community hospital and AMT), each ComprehENotes question was preprocessed by NoteAid and the results were embedded into the test web application directly. Definitions for terms were added to the web application as tooltip text. Terms were underlined to indicate that a definition was available, and when a participant hovered over a defined term, the definition would automatically display (Figure 2). This behavior was also described in the introductory text paragraph of the web application so that the participants were aware of the definitions and knew how to access them.

Data Collection

Community Participant Recruitment

With IRB approval, participant recruitment was conducted at LGH. Staff members of the research team (WH and MT) approached diabetes patients in the waiting room before or after their appointments. Patients and persons accompanying patients in the waiting room were eligible to participate if they were over 18 years old, able to speak and read English, and comfortable using a tablet. In some cases, patients who started the survey could not finish it before their appointments and therefore were not included in the study. Partial responses were not included in our analyses.

The staff approached a potential participant and asked whether he/she would be interested in a short online English survey to see how he/she understands medical jargon terms in doctors' notes. Then, the staff explained the fact sheet, including information regarding time to complete, IRB approval, and contact information. The staff also noted that the EHR notes in the questionnaire were not from their own personal EHR notes. We informed participants that once the survey started, they could discontinue participation at any time. Each participant was given a US $10 gift certificate. Participants were randomly assigned to the control or treatment group when they accessed the web page to complete the test via a random number generator implemented in Python.

AMT Participant Recruitment

A total of 200 participants were recruited on AMT (100 in the control group and 100 in the intervention group). Task visibility on the AMT platform was restricted to AMT workers located in the United States with prior task approval rates above 95%. The prior task approval rate percentage was used as an indicator of high-quality prior work, and selecting the United States as the location was used as a proxy for English proficiency. We did not collect any medical history information from participants. While the test administered to the AMT and community hospital participants was the same, there were several differences in test administration. First, AMT participants completed the task remotely, while the community hospital participants completed the task locally and under the supervision of research staff. Second, community hospital participants were randomly assigned to either the control or intervention group upon enrollment. On AMT, the task was implemented as a parallel randomized study. We first collected responses from 100 AMT participants for the control task (ie, no access to NoteAid). We then created and released a second AMT task with the intervention task. This intervention task was not available to AMT participants who participated in the control task. AMT participants were paid US $3 to complete the task. In pilot studies, we observed that AMT participants typically took between 15 and 20 minutes to complete the task; therefore, a US $3 payment approximated a US $9 to $12 hourly wage.

Data Analysis

For our specific hypotheses regarding the effects of NoteAid in the two participant recruitment sources (hereafter referred to as "*source*") (RQ1-3), we ran a two-way analysis of variance (ANOVA) to compare the four groups in our data set, using the proportion of the passage-item pairs answered correctly as the dependent variable and *source* (community hospital vs AMT) and *condition* (control vs intervention) as two crossed factors. Specifically, for our third research question, we tested the interaction effect between *source* and *condition*. An interaction would indicate that NoteAid's effect on the test score differs between AMT participants and community hospital participants. In this case, for our primary research question, we would compare community hospital participants in the control and intervention groups to determine the effectiveness of NoteAid among community hospital participants, and for our secondary research question, we would compare the community hospital participants and AMT participants separately under the intervention and control conditions. If the interaction is not significant, the two main effects would be tested to address the primary and secondary research questions.

XSL•FO
RenderX

To examine the effects of participant demographic characteristics and possible variations in the effect of NoteAid on the test score among different demographic categories (RQ4-5), we further considered a model with three-way interactions among *condition*, *source*, and each of the demographic variables (age, race, education, and gender) along with relevant lower order effects. The four effects concerning the same demographic variables were treated as one family, and each of them was tested at level α'=.0125. If an interaction was detected and simple effects were examined, their tests were further adjusted with Bonferroni correction. Pairwise comparisons among demographic categories were also adjusted with the Bonferroni method.

Given the discrete nature of the dependent variable and the likely ceiling effect due to the overall good performance, we supplemented the ANOVA with generalized linear models (GLMs), treating the dependent variable as a binomial outcome with possible overdispersion to account for individual differences.

Results

Demographics

We recruited a total of 188 participants at the community hospital location from the end of December 2019 to the beginning of March 2020. Results from 174 participants were included in the final analysis. Of the 174 participants, 141 were patients and 33 were persons accompanying patients. Fourteen participants were recruited and began the task, but did not complete it as they were called to their appointments and therefore were not included in our final analyses.

Characteristics of the AMT participants and the community hospital participants are presented in Table 1. The distribution of age was very different between the groups. The AMT participants were primarily younger, with the majority of AMT participants under 34 years old, while the majority of community hospital participants were over 55 years old. There were also differences in education. A majority of AMT participants had a bachelor's or master's degree, while fewer community hospital participants had either degree. The majority of community hospital participants had at the most an associate's degree.

Within the community hospital sample, the age, education, race, and gender profiles in the control group were similar to those in the intervention group, as expected for a randomized experiment. Chi-square independence tests for contingency tables were not significant after multiplicity adjustment. The same was true for the AMT sample.

XSL•FO
RenderX

Table 1. Demographic information of the study participants.

Characteristic	AMT[a], n (%)			Community hospital, n (%)			Overall, n (%)
	Baseline (n=100)	NoteAid (n=100)	Total (n=200)	Baseline (n=85)	NoteAid (n=89)	Total (n=174)	Total (N=374)
Age							
18-21	2 (2.0%)	1 (1.0%)	3 (1.5%)	2 (2.4%)	3 (3.4%)	5 (2.9%)	8 (2.1%)
21-34	54 (54.0%)	61 (61.0%)	115 (57.5%)	16 (18.8%)	7 (7.9%)	23 (13.2%)	138 (36.9%)
35-44	27 (27.0%)	25 (25.0%)	52 (26.0%)	12 (14.1%)	8 (9.0%)	20 (11.5%)	72 (19.3%)
45-54	11 (11.0%)	9 (9.0%)	20 (10.0%)	12 (14.1%)	18 (20.2%)	30 (17.2%)	50 (13.4%)
55-64	5 (5.0%)	3 (3.0%)	8 (4.0%)	14 (16.5%)	30 (33.7%)	44 (25.3%)	52 (13.9%)
≥65	1 (1.0%)	1 (1.0%)	2 (1.0%)	27 (31.8%)	18 (20.2%)	45 (25.9%)	47 (12.6%)
Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.4%)	5 (5.6%)	7 (4.0%)	7 (1.9%)
Education							
Less than high school	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (4.7%)	6 (6.7%)	10 (5.7%)	10 (2.7%)
High school	24 (24.0%)	28 (28.0%)	52 (26.0%)	34 (40.0%)	34 (38.2%)	68 (39.1%)	120 (32.1%)
Associate's degree	17 (17.0%)	22 (22.0%)	39 (19.5%)	15 (17.6%)	14 (15.7%)	29 (16.7%)	68 (18.2%)
Bachelor's degree	53 (53.0%)	42 (42.0%)	95 (47.5%)	18 (21.2%)	16 (18%)	34 (19.5%)	129 (34.5%)
Master's degree	6 (6.0%)	8 (8.0%)	14 (7.0%)	12 (14.1%)	14 (15.7%)	26 (14.9%)	40 (10.7%)
Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.4%)	5 (5.6%)	7 (4.0%)	7 (1.9%)
Race							
African American	8 (8.0%)	9 (9.0%)	17 (8.5%)	3 (3.5%)	2 (2.2%)	5 (2.9%)	22 (5.9%)
American Indian	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.1%)	1 (0.6%)	1 (0.3%)
Asian	8 (8.0%)	2 (2.0%)	10 (5.0%)	8 (9.4%)	10 (11.2%)	18 (10.3%)	28 (7.5%)
Hispanic	15 (15.0%)	6 (6.0%)	21 (10.5%)	11 (12.9%)	14 (15.7%)	25 (14.4%)	46 (12.3%)
White	69 (69.0%)	83 (83.0%)	152 (76.0%)	61 (71.8%)	57 (64.0%)	118 (67.8%)	270 (72.2%)
Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.4%)	5 (5.6%)	7 (4.0%)	7 (1.9%)
Gender							
Female	39 (39.0%)	42 (42.0%)	81 (40.5%)	44 (51.8%)	43 (48.3%)	87 (50.0%)	168 (44.9%)
Male	61 (61.0%)	58 (58.0%)	119 (59.5%)	38 (44.7%)	40 (44.9%)	78 (44.8%)	197 (52.7%)
Refrain	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.2%)	1 (1.1%)	2 (1.1%)	2 (0.5%)
Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (2.4%)	5 (5.6%)	7 (4.0%)	7 (1.9%)

[a]AMT: Amazon Mechanical Turk.

Effect of the Intervention and Participant Recruitment Source

Table 2 shows the descriptive statistics (proportion correct) of the four groups. Results of the two-way ANOVA are presented in Table 3. Note the CIs in Table 3 are not simultaneous CIs, but one-at-a-time CIs.

The interaction effect in the ANOVA was not significant (P=.89), suggesting no evidence that the effect of NoteAid was different across the two participant recruitment sources. Further analyses show that both the main effects of *source* (AMT or community hospital) and of *condition* (baseline or intervention) were significant. Participants who took the ComprehENotes test on AMT on average scored higher than the community hospital participants, and the difference was significant. In addition, participants who had access to NoteAid scored higher than those who did not have access to NoteAid. The difference was again significant.

Our analysis with GLMs yielded similar results. The main effect of *condition* was significant (odds ratio [OR] 1.23, 95% CI 1.10-1.38; P<.001), the main effect of *source* was significant (OR 1.33, 95% CI 1.18-1.49; P<.001), and the interaction effect between *source* and *condition* was not significant (P=.72).

XSL•FO
RenderX

Table 2. Summary statistics for the source by condition contingency table in our analysis of variance.

Source	Condition	
	Control[a]	Intervention[a]
AMT[b]	0.756 (0.246), n=100	0.830 (0.201), n=100
Community hospital	0.646 (0.179), n=85	0.727 (0.191), n=89

[a]Data are presented as mean proportion correct (SD), sample size.

[b]AMT: Amazon Mechanical Turk.

Table 3. Analysis of variance table.

Variable	df	Sum squares	Mean squares	*F*	Cohen *d* (95% CI)
Source	1	1.06	1.06	24.70[a]	0.52 (0.31 to 0.72)
Condition	1	0.56	0.56	13.06[a]	0.37 (0.17 to 0.58)
Source × condition	1	0.001	0.001	0.02	0.03 (−0.38 to 0.44)
Residuals	370	15.88	0.04	N/A[b]	N/A

[a]*P*<.001.

[b]N/A: not applicable.

Effects of Demographic Variables

To study the effects of demographic variables, the single case of an American Indian, the two cases where individuals refrained from reporting gender, and the seven cases with missing demographic information were removed from the data set before analyses. To mitigate data sparsity in the contingency table, the category “less than high school” was combined with “high school” in *education*, and the highest and lowest age groups were combined with their adjacent groups. Results from ANOVA and GLMs gave qualitatively the same results. However, ANOVA yielded predicted scores exceeding one and demonstrated clear violation of homoscedasticity. We chose to report results from the GLMs. An outlying residual was identified in the analyses, but its removal yielded similar results.

The only significant effect involving a demographic variable was the interaction between *source* and *education* (χ^2_3=16.9, *P*<.001). This shows that the effect of *education* differed across the AMT and community hospital participants. Separate analyses of data from the two groups found that *education* did not have a significant main effect (χ^2_3=2.05, *P*=.56) in the AMT group, but had a significant main effect (χ^2_3=37.30, *P*<.001) in the community hospital group. Pairwise comparisons revealed that community hospital participants with “high school or lower” education had significantly lower performance than those with a bachelor’s degree and those with a master’s degree. The interaction between *condition* and every demographic variable was not significant, suggesting no evidence of any variations in the effect of NoteAid for people in different demographic categories.

Discussion

Principal Results

In this work, we have demonstrated the effectiveness of NoteAid for improving EHR note comprehension in participants recruited from two different sources (a community hospital setting and AMT). For both samples, access to NoteAid significantly improved ComprehENotes scores (RQ1). To the best of our knowledge, NoteAid is the only tool available that has been shown to improve patient health literacy, both in this work and in prior work [25,26]. Samples recruited from these two sources varied, in particular with regard to age and education. AMT participants were younger and more educated than participants recruited from the community hospital. Consistent with prior findings on health literacy and education levels [7], the community hospital participants scored significantly lower than the AMT participants on the ComprehENotes test (RQ2).

Although there were demographic differences between the participant recruitment sources (LGH and AMT), we did not find evidence that the effect of NoteAid was different across the recruitment sources (RQ3). We found that age, race, and gender did not have a significant effect on scores. We did find that the effect of education differed between the two sources (RQ4). For the LGH sample, participants with a high school education or less had significantly lower scores than individuals with a bachelor’s or master’s degree. This result is consistent with existing literature that those with less education are at greater risk for low health literacy [7]. In contrast, no education effect was detected for the online platform AMT participants (RQ5). It is important to include participants recruited from an actual hospital setting to confirm the effectiveness of health literacy tests (eg, ComprehENotes) and tools (eg, NoteAid) across samples.

Results from both LGH and AMT confirmed prior research on the effectiveness of NoteAid for improving participant EHR

XSL•FO
RenderX

note comprehension [18,26]. We showed that participants with access to NoteAid achieved higher scores on average than those without access to NoteAid, consistent with both patient self-reporting and empirical analyses in prior work [18,26]. However, this is the first study to test NoteAid empirically in participants recruited from a community hospital setting. While the effectiveness of NoteAid was consistent across our participant recruitment sources, participants from the two sources varied in terms of key characteristics (eg, age and education).

Electronic patient portals are becoming more common, with as many as 44 million patients estimated to have access to their notes [32]. Notes are being made available as part of the OpenNotes initiative [33] and in a variety of other health care settings, from the US Department of Veterans Administration [34,35] to private organizations [36]. As a result, more information about personal health is available to patients. In a recent study, surveyed patients reported that they understood "most or all" of the content in their EHR notes [37]. However, all patients surveyed had previously read at least one of their notes, and most would be considered highly activated patients. Patients who can and do currently view and understand their notes typically have high patient activation [3,38]. They are active in managing their health care, have lower likelihood of outcomes, such as emergency department visits, and are more likely to engage in preventive care [3]. Individuals with low patient activation typically do not understand their role in the health care process and are less likely to participate in tasks associated with health care management (eg, viewing their patient portal) [38]. For those surveyed patients with a high school diploma or lower, self-reported note understanding was significantly lower, consistent with our results [37]. In addition, a common comment from survey respondents was that they had difficulty understanding medical jargon in the note, with respondents requesting access to jargon definitions. Self-reported levels of understanding may not reflect actual understanding, as measured by validated instruments such as ComprehENotes [39,40]. Our results showed significant improvement in comprehension regardless of education level, indicating that NoteAid is an effective intervention.

There are two key challenges for ensuring that patients can realize the benefits associated with accessing their own notes as follows: (1) defining medical jargon so that patients understand the content of their notes, and (2) motivating low activation patients to view their notes and take an active role in managing their health care. The NoteAid tool directly addresses the first challenge, as demonstrated by this work and prior work. As more patients have access to their notes, providing access to definitions at the same time can reduce issues with patients not understanding the content in their newly available notes. It may also indirectly address motivation. If low activation patients struggle with medical jargon, having the jargon automatically defined can reduce the barrier to entry for their participation in care. Implementation of NoteAid directly into EHR software, for example, via an application programming interface (API), would allow for patients to have jargon terms defined within the patient portal itself, without needing to search for definitions outside of the platform. This aligns with a recent call for "easy-to-understand information" as part of a proposal for improving patient portals [41]. Further, NoteAid can be used by anyone who assists in the management of a patient's care. If a patient chooses to share an EHR note with a family member, the family member can use NoteAid to define medical jargon terms, so that he or she can better understand the note and better assist with the patient's care. In this case, even patients who themselves might struggle to use NoteAid or read their notes can benefit from the tool.

Limitations

There are several limitations in this study. First, local recruitment of individuals for our trial had to be halted due to COVID-19. While we were able to gather 174 responses from local participants and identify a significant effect of using NoteAid, a larger sample size would allow for more fine-grained analysis of the results, for example, examination of scores according to various demographic characteristics. While this might be seen as an argument in favor of moving such data collection efforts to crowdsourcing platforms, our results indicate that local population scores are significantly lower than the scores of users on crowdsourcing platforms (RQ2). In addition, the demographic differences between individuals in the local population and individuals on AMT indicate that simply relying on AMT is not sufficient. We confirm prior work demonstrating the effect of NoteAid on EHR note comprehension in AMT workers [26], but go one step further to show that the tool is useful for local patients (a considerably different cohort).

A second limitation is with regard to the NoteAid tool. Certain terms or acronyms in EHR notes can have more than one meaning. The task of correctly identifying the appropriate meaning of an ambiguous term is a well-studied problem in natural language processing called *word sense disambiguation*. While a number of methods have been proposed to handle this problem [42-44], there is active research to improve these models. A NoteAid system in a production environment would need to be able to disambiguate between possible definitions so that a patient would receive the definition for the correct term in his or her own note [25]. In the case of our study, NoteAid definitions were manually inspected and added to the ComprehENotes test as part of our web application implementation. Therefore, we were able to confirm the correct definitions for the terms in our EHR note snippets before running our tasks.

Finally, our local subject group was restricted to diabetes patients and those accompanying them to their appointments. There is a risk that this local setting is too narrow in terms of scope for the results to be more generally applicable. However, the ComprehENotes test is not a test of diabetes EHR note comprehension. The test includes questions from EHR notes related to a number of diseases, such as diabetes, cancer, heart failure, and hypertension, in order to include a wide range of subjects in the assessment [24]. Coupled with the fact that we saw similar results in the AMT participants, we believe that the results could be generalized beyond diabetes patients to a wider patient cohort.

XSL•FO
RenderX

Conclusion and Future Work

The findings reported here provide evidence of the effectiveness of NoteAid for improving EHR note comprehension across two different participant samples: patients from a community hospital and participants from a popular online crowdsourcing platform. Despite the demographic and education differences between the two samples, NoteAid improved scores on the ComprehENotes test for both, indicating that it is an effective intervention for improving EHR note comprehension. These results support broader use of ComprehENotes as an EHR note comprehension test and NoteAid as an effective tool for improving EHR note understanding.

Future work should explore patient personalization in NoteAid. Providing lay definitions for all medical jargon in a note may lead to information overload for patients with low education and may be unnecessary for patients with high education. To improve comprehension for low health literacy patients, other mediums (eg, short animations and text definitions) may be more effective. Operationally, making NoteAid jargon definitions available for a patient's own notes via an API would increase the ways in which researchers and EHR vendors can implement jargon definitions so that patients have access to them.

Future work applying the NoteAid tool to other contexts is another interesting direction. While NoteAid has been shown to improve patient EHR note comprehension [25], the methodology of the tool is generalizable. The concept identification-definition linking process can be applied to other texts where complex jargon is common to improve readability and understanding. If a lay-language dictionary is built for a particular domain (eg, legal text), it would be possible to use NoteAid to identify and define those complex terms so that individuals can more easily understand the text.

Acknowledgments

The authors thank the anonymous Amazon Mechanical Turk participants who participated in the Amazon Mechanical Turk tasks and the subjects who participated in the community hospital tasks. The authors would like to thank Harmon Jordan for his helpful conversations with regard to this work. This work was supported in part by grant I01HX001457 from the Health Services Research & Development, US Department of Veterans Affairs and by grant R01LM012817 from the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not represent the views of the NIH or the US Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 985 KB - jmir_v23i5e26354_app1.pdf]

References

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC: The National Academies Press; 2001.
2. Cliff B. The evolution of patient-centered care. J Healthc Manag 2012;57(2):86-88. [Medline: 22530289]
3. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
4. Begum N, Donald M, Ozolins IZ, Dower J. Hospital admissions, emergency department utilisation and patient activation for self-management among people with diabetes. Diabetes Res Clin Pract 2011 Aug;93(2):260-267. [doi: 10.1016/j.diabres.2011.05.031] [Medline: 21684030]
5. 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]
6. Diviani N, van den Putte B, Giani S, van Weert JC. 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]
7. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. National Center for Education Statistics. 2006. URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483# [accessed 2021-05-06]
8. Bass PF, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. Acad Med 2002 Oct;77(10):1039-1041. [doi: 10.1097/00001888-200210000-00021] [Medline: 12377684]
9. Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P. The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. Am J Obstet Gynecol 2002 May;186(5):938-943. [doi: 10.1067/mob.2002.122091] [Medline: 12015518]
10. Vastag B. Low health literacy called a major problem. JAMA 2004 May 12;291(18):2181-2182. [doi: 10.1001/jama.291.18.2181] [Medline: 15138230]

XSL•FO
RenderX

11. Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 2006 Aug;21(8):806-812 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00539.x] [Medline: 16881938]
12. Schapira MM, Fletcher KE, Hayes A, Eastwood D, Patterson L, Ertl K, et al. The development and validation of the hypertension evaluation of lifestyle and management knowledge scale. J Clin Hypertens (Greenwich) 2012 Jul;14(7):461-466 [FREE Full text] [doi: 10.1111/j.1751-7176.2012.00619.x] [Medline: 22747619]
13. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6):557-566. [doi: 10.1002/pon.673] [Medline: 12923796]
14. Lerner EB, Jehle DV, Janicke DM, Moscati RM. Medical communication: do our patients understand? Am J Emerg Med 2000 Nov;18(7):764-766. [doi: 10.1053/ajem.2000.18040] [Medline: 11103725]
15. Reading SR, Go AS, Fang MC, Singer DE, Liu IA, Black MH, AnticoagulationRisk Factors in Atrial Fibrillation–Cardiovascular Research Network (ATRIA‐CVRN) Investigators. Health Literacy and Awareness of Atrial Fibrillation. J Am Heart Assoc 2017 Apr 11;6(4):1-10 [FREE Full text] [doi: 10.1161/JAHA.116.005128] [Medline: 28400367]
16. Halbach SM, Enders A, Kowalski C, Pförtner TK, Pfaff H, Wesselmann S, et al. Health literacy and fear of cancer progression in elderly women newly diagnosed with breast cancer--A longitudinal analysis. Patient Educ Couns 2016 May;99(5):855-862. [doi: 10.1016/j.pec.2015.12.012] [Medline: 26742608]
17. Jeppesen KM, Coyle JD, Miser WF. Screening questions to predict limited health literacy: a cross-sectional study of patients with diabetes mellitus. Ann Fam Med 2009 Jan 01;7(1):24-31 [FREE Full text] [doi: 10.1370/afm.919] [Medline: 19139446]
18. Polepalli Ramesh B, Houston T, Brandt C, Fang H, Yu H. Improving patients' electronic health record comprehension with NoteAid. Stud Health Technol Inform 2013;192:714-718. [Medline: 23920650]
19. Ylitalo KR, Meyer MRU, Lanning BA, During C, Laschober R, Griggs JO. Simple screening tools to identify limited health literacy in a low-income patient population. Medicine (Baltimore) 2018 Mar;97(10):e0110 [FREE Full text] [doi: 10.1097/MD.0000000000010110] [Medline: 29517689]
20. Zheng J. Computational Approaches to Assisting Patients' Medical Comprehension from Electronic Health Records. University of Massachusetts Amherst. 2020. URL: https://scholarworks.umass.edu/dissertations_2/1972/ [accessed 2021-05-06]
21. Noblin AM, Wan T, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
22. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
23. Walker J, Meltsner M, Delbanco T. US experience with doctors and patients sharing clinical notes. BMJ 2015 Feb 10;350:g7785. [doi: 10.1136/bmj.g7785] [Medline: 25670182]
24. Lalor JP, Wu H, Chen L, Mazor KM, Yu H. ComprehENotes, an Instrument to Assess Patient Reading Comprehension of Electronic Health Record Notes: Development and Validation. J Med Internet Res 2018 Apr 25;20(4):e139 [FREE Full text] [doi: 10.2196/jmir.9380] [Medline: 29695372]
25. Chen J, Druhl E, Polepalli Ramesh B, Houston TK, Brandt CA, Zulman DM, et al. A Natural Language Processing System That Links Medical Terms in Electronic Health Record Notes to Lay Definitions: System Development Using Physician Reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]
26. Lalor JP, Woolf B, Yu H. Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers. J Med Internet Res 2019 Jan 16;21(1):e10793 [FREE Full text] [doi: 10.2196/10793] [Medline: 30664453]
27. Ross J, Irani L, Silberman MS, Zaldivar A, Tomlinson B. Who are the crowdworkers? Shifting demographics in Mechanical Turk. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. 2010 Presented at: CHI '10: CHI Conference on Human Factors in Computing Systems; April 2010; Atlanta, GA, USA p. 2863-2872. [doi: 10.1145/1753846.1753873]
28. Difallah D, Filatova E, Ipeirotis P. Demographics and Dynamics of Mechanical Turk Workers. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: Eleventh ACM International Conference on Web Search and Data Mining; February 2018; Marina Del Rey, CA, USA p. 135-143. [doi: 10.1145/3159652.3159661]
29. Royer JM, Hastings CN, Hook C. A Sentence Verification Technique for Measuring Reading Comprehension. Journal of Reading Behavior 2016 Sep 11;11(4):355-363. [doi: 10.1080/10862967909547341]
30. Baker FB, Kim SH. Item Response Theory: Parameter Estimation Techniques, Second Edition. Boca Raton, FL: CRC Press; 2004.
31. Mellis AM, Bickel WK. Mechanical Turk data collection in addiction research: utility, concerns and best practices. Addiction 2020 Oct 24;115(10):1960-1968. [doi: 10.1111/add.15032] [Medline: 32135574]

XSL•FO
RenderX

32. DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al. The Views and Experiences of Clinicians Sharing Medical Record Notes With Patients. JAMA Netw Open 2020 Mar 02;3(3):e201753 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1753] [Medline: 32219406]
33. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
34. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
35. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
36. Huang JS, Yueh R, Ma S, Cruz R, Bauman L, Choi LJ. Adolescents' and Young Adults' Satisfaction with and Understanding of Medical Notes from a Pediatric Gastroenterology Practice: A Cross-Sectional Cohort Study. J Pediatr 2019 Dec;215:264-266. [doi: 10.1016/j.jpeds.2019.06.052] [Medline: 31377044]
37. Leveille SG, Fitzgerald P, Harcourt K, Dong Z, Bell S, O'Neill S, et al. Patients Evaluate Visit Notes Written by Their Clinicians: a Mixed Methods Investigation. J Gen Intern Med 2020 Dec 15;35(12):3510-3516. [doi: 10.1007/s11606-020-06014-7] [Medline: 32671721]
38. Hibbard JH. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017 Jan;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
39. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, 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 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
40. Snibsøer AK, Ciliska D, Yost J, Graverholt B, Nortvedt MW, Riise T, et al. Self-reported and objectively assessed knowledge of evidence-based practice terminology among healthcare students: A cross-sectional study. PLoS One 2018 Jul 12;13(7):e0200313 [FREE Full text] [doi: 10.1371/journal.pone.0200313] [Medline: 30001380]
41. DesRoches CM, Walker J, Delbanco T. Care Partners and Patient Portals-Faulty Access, Threats to Privacy, and Ample Opportunity. JAMA Intern Med 2020 Jun 01;180(6):850-851. [doi: 10.1001/jamainternmed.2020.0514] [Medline: 32364559]
42. Savova GK, Coden AR, Sominsky IL, Johnson R, Ogren PV, de Groen PC, et al. Word sense disambiguation across two domains: biomedical literature and clinical notes. J Biomed Inform 2008 Dec;41(6):1088-1100 [FREE Full text] [doi: 10.1016/j.jbi.2008.02.003] [Medline: 18375190]
43. Stevenson M, Guo Y. Disambiguation of ambiguous biomedical terms using examples generated from the UMLS Metathesaurus. J Biomed Inform 2010 Oct;43(5):762-773 [FREE Full text] [doi: 10.1016/j.jbi.2010.06.001] [Medline: 20541624]
44. Wu Y, Xu J, Zhang Y, Xu H. Clinical Abbreviation Disambiguation Using Neural Word Embeddings. In: Proceedings of BioNLP 2015. 2015 Presented at: BioNLP 2015; July 2015; Beijing, China p. 171-176. [doi: 10.18653/v1/w15-3822]

Abbreviations

AMT: Amazon Mechanical Turk
ANOVA: analysis of variance
API: application programming interface
EHR: electronic health record
GLM: generalized linear model
IRB: Institutional Review Board
IRT: item response theory
LGH: Lowell General Hospital
NLP: natural language processing
OR: odds ratio

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.12.20; peer-reviewed by H Imeri, D Cebo; comments to author 04.01.21; revised version received 07.01.21; accepted 06.04.21; published 13.05.21.

Please cite as:
Lalor JP, Hu W, Tran M, Wu H, Mazor KM, Yu H
Evaluating the Effectiveness of NoteAid in a Community Hospital Setting: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Patients
J Med Internet Res 2021;23(5):e26354
URL: https://www.jmir.org/2021/5/e26354
doi:10.2196/26354
PMID:33983124

©John P Lalor, Wen Hu, Matthew Tran, Hao Wu, Kathleen M Mazor, Hong Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.05.2021. 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

Patient Rationales Against the Use of Patient-Accessible Electronic Health Records: Qualitative Study

Hanne Støre Valeur[1], MD; Anne Kveim Lie[1], MD, PhD; Kåre Moen[1], MD, PhD

Institute of Health and Society, University of Oslo, Oslo, Norway

Corresponding Author:
Hanne Støre Valeur, MD
Institute of Health and Society
University of Oslo
Postboks 1130 Blindern
Oslo, 0318
Norway
Phone: 47 45248693
Email: h.s.valeur@medisin.uio.no

Abstract

Background: Patient-accessible electronic health records (PAEHRs) enable patients to access their health records through a secure connection over the internet. Although previous studies of patient experiences with this kind of service have shown that a majority of users are positive toward PAEHRs, little is known about why some patients occasionally or regularly choose not to use them. A better understanding of why patients may choose not to make use of digital health services such as PAEHRs is important for further development and implementation of services aimed at having patients participate in digital health services.

Objective: The objective of the study was to explore patients' rationales for not embracing online access to health records.

Methods: Qualitative interviews were conducted with 40 patients in a department of internal medicine in a Norwegian hospital in 2018-2019. Interview transcripts were subjected to thematic content analysis. In this paper, we focus on the subject of nonuse of PAEHRs.

Results: We identified 8 different rationales that study participants had for not embracing PAEHRs. When patients reflected on why they might not use PAEHRs, they variously explained that they found PAEHRs unnecessary (they did not feel they were useful), impersonal (they preferred oral dialogue with their doctor or nurse over written information), incomprehensible (the records contained medical terminology and explanations that were hard to understand), misery oriented (the records solely focused on disease), fear provoking (reading the records could cause unwanted emotional reactions), energy demanding (making sense of the records added to the work of being a patient), cumbersome (especially among patients who felt they did not have the necessary digital competence), and impoverishing (they were skeptical about the digital transformation of individual and social life).

Conclusions: It is often assumed that the barriers to PAEHR use are mostly practical (such as lack of hardware and access to the internet). In this study, we showed that patients may have many other valid reasons for not wanting to adopt this kind of service. The results can help guide how PAEHRs and other digital health services are promoted and presented to patients, and they may suggest that the goal of a given new digital health service should not necessarily be full uptake by all patients. Rather, one should recognize that different patients might prefer and benefit from different kinds of "analog" and digital health services.

(J Med Internet Res 2021;23(5):e24090) doi:10.2196/24090

KEYWORDS

patient-accessible electronic health records; open notes; active patients; patients' perspective; patient portal; electronic health records; participation

XSL•FO
RenderX

Introduction

Background

Over the past decade, patient-accessible electronic health records (PAEHRs) have been requested, developed, introduced, and advocated for in health care systems in several nations [1,2]. The existing literature conveys the impression that a majority of patients are satisfied with PAEHRs and argue that they can increase patient involvement, improve patient-provider communication, support self-management, and promote patient empowerment [3-6]. In a survey of patient experiences in Norway, the majority of users were satisfied overall with PAEHRs [7] and perceived online access to their medical records as useful for tasks such as looking up health information, keeping track of ongoing treatments, preparing for medical appointments, and sharing medical documentation with others. Similar findings have been reported from other countries [8-10].

Notwithstanding this, there are always significant minorities of patients in PAEHR surveys who are less enthusiastic about PAEHRs. For example, 6% of respondents in a Swedish survey indicated that they did not think PAEHRs led to better understanding [8], 12% of respondents in a US survey were concerned about their privacy in connection with PAEHR use [9], and 19% of respondents in the Norwegian study mentioned above had become worried because of information they found in their health records [7].

Little is known about the perspectives of patients who do not access their records. A qualitative study from Sweden included 15 patients with cancer without experience using PAEHRs [11]. The main argument of these patients for not accessing records was reported to be "that they have a good relationship with their physician and that they receive the information they need," but their rationales for not using the service were not explored further. A better understanding of why patients may resist using new digital health services like PAEHRs is important for further successful development and implementation of digital services.

Digitalization in Norway

Compared with many other countries, Norway has high coverage of internet access and use, with 98% of the population aged between 16 and 79 years having used the internet for different purposes during the last 3 months [12]. A range of different public and commercial services have been extensively digitalized (eg, banking—to the degree that it is now highly inconvenient to handle personal banking outside the web), and most citizens employ digital tools in their everyday lives [13]. Therefore, we saw Norway as a good setting for studying patients' views on electronic health services that go beyond issues concerning the digital divide.

Providing all patients with electronic access to their health records has been a goal for the Norwegian government since 2012 [14]. A white paper entitled "En innbygger, en journal" ("One citizen, one health record") argued that such access "gives citizens overview of their own information and knowledge about their health and illness. This provides a basis for co-decision making and active cooperation between health care professionals and patients and clients. Patients and clients will, through access and participation, be able to point out errors in the health records and improve the quality of the documentation" [14].

"Pasientjournal": The PAEHR Used in Norwegian Specialized Health Care

In December 2015, PAEHRs were introduced under the name Pasientjournal by the publicly owned Northern Norway Regional Health Authority, the owner of the hospital in which fieldwork for this study was carried out. The stated goal was to "increase patient empowerment and patients' involvement in their own health, and improve the quality of services" [15]. In a video promoting the service, the Pasientjournal is described as a tool for becoming an "enlightened patient" [16].

The Pasientjournal service is now available to anyone seeking public specialist health care in 3 of Norway's 4 health regions, which cover roughly 80% of the Norwegian population. It gives patients access to almost all documents in their medical records. They can retrieve and read doctors' notes, discharge notes, outpatient clinic notes, nurses' documentation, physiotherapy notes, referral notes, radiology reports, and some pathology test results. Laboratory results are not yet available but will be in the near future. All documents are accessible by the patient as soon as they have been digitally signed by the responsible health care professional. The list of documents is ordered chronologically and can comprise a few entries to hundreds of entries depending on how much contact the patient has had with health care services. Patients can click on any given document to read its full contents. Apart from reading, the patient cannot take any other action online. If one has a question or discovers something incorrect, one is encouraged to contact the hospital for further guidance. Patients log onto the service using the same identification procedure as for several other public digital services.

Based on qualitative research in a Norwegian hospital, the aim of this paper was to explore what patients conceive of as the main weaknesses and disadvantages of PAEHRs as well as their rationales for not actively embracing online access to their health records. We based our analysis on interviews with PAEHR users (both regular and occasional) as well as with patients who had never accessed their records online.

Methods

The fieldwork for this study was carried out in 2018-2019 in a hospital owned by the Northern Norway Regional Health Authority. We have decided to conceal the name of the hospital as part of our efforts to protect study participants' personal information.

Study Design

This paper draws on a study that was inspired by the ethnographic research tradition. The first author (HSV) did fieldwork for a period of 12 weeks in the hospital's department of internal medicine, which consists of several wards and outpatient clinics and covers a range of medical specialties (nephrology, endocrinology, cardiology, pulmonology, gastroenterology, hematology, and oncology). Fieldwork entailed participant observation as well as semistructured

XSL•FO
RenderX

interviewing of patients and health care professionals, with the overarching goal of exploring how PAEHRs are incorporated into everyday practices in a hospital. This paper is based on an analysis of the patient interviews conducted in the study.

Qualitative Interviews and Recruitment of Study Participants

A total of 40 patients were interviewed. They were recruited in the course of participant observation, during which the first author (HSV) accompanied doctors and nurses while they were doing their daily work in the department. Patients who gave their consent were invited to take part in an interview. The researcher had no knowledge of the invited patients' experience with PAEHRs prior to inviting them to participate in the study, unless this had become evident in the observed interaction with their doctor or nurse. To help ensure that patients were competent to give consent and in a condition that allowed them to sit through an interview, potential study participants were discussed with their care providers before they were invited to participate in the study. All but one of the patients who were asked to participate accepted the invitation. They received information about the study both orally and in writing before they were asked to provide written consent to participate.

Patients were recruited as the first author moved around the clinic and aimed to include patients with a variety of ages, genders, and diagnoses. The variety was obtained by recruiting patients from both the inpatient and the outpatient clinic and by following health care professionals from different medical specialties. Of the 40 interviewees, 22 were women and 18 were men. Their ages varied between 21 and 84 years. The study participants had a range of different medical conditions in the domains of nephrology, cardiology, cancer, hematology, lung diseases, and infections. Among the 40 interviewees, 32 knew about Pasientjournal prior to the interview and 8 did not. Among the latter, 6 indicated that they felt positive about the idea of logging into and reading their medical records while 2 did not. A little less than half (ie, 16 individuals) had used PAEHRs either occasionally or regularly prior to the interview.

The interviews, which lasted between 15 and 60 minutes, took place by the hospital bed in the case of admitted patients and in offices and examination rooms in the case of outpatients. An interview guide listing topics to be covered (eg, the patient's history of illness, prior knowledge and use of the PAEHR, health information–seeking habits, and other digital habits) was used as a memory aid during the conversations, but participants were encouraged to speak freely and the interviewer sought to let the participants take the lead in the dialogue. The interviewer's aim was to be a conversation partner who traveled along with the patient into their stories rather than someone trying to "dig out" information from them (this style of interviewing was inspired by Brinkmann and Kvale's metaphor of the interviewer as a "traveler" [17]).

In this paper, we draw on the data generated through qualitative interviewing, focusing specifically on patients' rationales for not using PAEHRs. We plan to return to other study findings in future works.

Data Analysis

All interviews were transcribed verbatim, and transcripts were read carefully several times. Following a stepwise deductive-inductive approach [18], we then coded the interviews empirically close with in vivo coding, meaning that many of the codes contained words and expressions used by the study participants themselves. The initial coding was done by the first author and then discussed thoroughly with the coauthors. In a subsequent step, codes were grouped into themes. Identified themes were discussed and reordered during several analysis meetings involving all authors until a common understanding had been reached. For the purpose of this paper, themes relating to why patients do or do not use PAEHRs were chosen and further analyzed. The themes identified in this article were generated on the basis of interviews with patients who did not want to use the PAEHR (in spite of knowledge of the service) as well as with patients who had personal experience with Pasientjournal.

Researcher Positions

Fieldwork was conducted by the first author (HSV). She is a woman in her 30s who is a trained physician and has clinical experience from work in hospitals (but she had not worked in a clinical setting for several years when fieldwork started and had not worked in a clinical setting after PAEHRs were introduced). She also has personal experience as a close relative of severely ill patients. Her medical background made it easy to move around in the clinic, as many hospital "codes" and routines were well known to her. She made efforts to explain to study participants that she was there as a researcher and not as a representative of the hospital management. At the hospital's request, she dressed in white scrubs during fieldwork. She thereby blended in with the staff and at times had to underline for the study participants that she was not present by virtue of her medical skills.

The second and third authors (AKL and KM, respectively) are also trained physicians, but both currently work full time as associate professors (in the fields of medical anthropology and medical history, respectively).

All authors engage with many digital tools in their personal and work lives and view themselves as highly computer literate and often appreciate digital solutions. None of the authors had clear opinions on PAEHRs prior to the study. They entered the study with curiosity and a wish to better understand patient experiences with PAEHR.

Ethical Considerations

The study was approved by the Norwegian Data Protection Authority (case no. 56987). Exemption from health care workers' duty to maintain confidentiality was obtained from the Regional Committee for Medical and Health Research. In the hospital, fieldwork was approved by the head of the clinic, the chief of research, and the data protection officer. All study participants gave informed written consent to participate and were told that they could withdraw from the study at any time without giving a reason. Directly person-identifying information were omitted in transcripts, and audio recordings were deleted after transcription. A key that linked the data to participants'

XSL•FO
RenderX

names and contact information was kept in a separate and securely locked location. All data were stored and processed in "Service for Sensitive Data" (Tjeneste for Sensitive Data), University of Oslo's platform for collecting, storing, analyzing, and sharing sensitive data in compliance with the Norwegian privacy regulations.

Results

The study participants had a range of different rationales both for and against the use of Pasientjournal. In this paper, our focus was on the latter, and in this section, we present the 8 different rationales against the use of Pasientjournal that we identified in the course of the study.

Unnecessary: No Need to Access the Health Records

One rationale against the use of PAEHRs was that the digital health records seemed unnecessary. For some of the study participants who had never accessed their PAEHR, the question was clearly not why they did not use the service but rather why they would ever do such a thing. They struggled to see how they could benefit from accessing Pasientjournal. Many felt they got the information they needed from their doctors and trusted them to provide necessary guidance and information. Some questioned how reading the records would make a difference for their health. For example, a woman in her 70s stated the following when the first author tried to dig into the reason behind her lack of interest in the PAEHR: "No, I do not believe that it would make me healthier" [Patient 12, a woman in her 70s; no PAEHR experience].

On the ward, paper copies of discharge notes were routinely given to all patients at departure. Many study participants felt that this was all the written information they needed:

> *"I don't know, really. The times I've been admitted, I've usually gotten the discharge note handed out, so I have not, like, had any reason to ... or felt a need for it. I have been given good information from the doctors I have been seeing as well"* [Patient 22, a woman in her 50s; no PAEHR experience]

In summary, some study participants had not felt motivated to spend time accessing the PAEHR, and they often argued that they had felt satisfied with the information they had been provided orally and in written discharge notes.

Impersonal: Prefer Dialogue to Information

Some study participants not only thought that information received in dialogue with health care workers was "good enough" but reasoned that it was superior to the information accessible through PAEHRs. Patient 14, for example, pointed out that medical records do not allow you to ask anyone the questions that pop up while you are reading and for that reason saw PAEHRs as inferior to face-to-face communication:

> *"Though I probably would get the same information, but you do not get to ask the questions you might have there and then, or get it explained if there is something you find hard to understand. I do think it is better to speak to someone about it than just reading it!"* [Patient 14, a woman in her 40s; PAEHR user]

A bit worried, she went on to ask if there was an intention to replace physical meetings with this form of written information, and added, "If so, I would see that as a loss."

For some, the thought of logging onto a computer to access information about one's own health seemed rather soulless. For Patient 11, for example, the human contact in face-to-face communication had considerable value:

> *"I'm a very verbal kind of person. If there is something that really annoys me, at work for example; the times where you are expected to find all the information for yourself, instead of just meeting up in the hallway and I can say: Let's meet up after lunch and discuss this. So I really like that human contact, and sometimes it's kind of a principle for me to pursue that line."* [Patient 11, a man in his 40s; PAEHR user]

A common finding throughout the fieldwork was that patients' use of superlatives in conversations about health care was typically related to individuals who had made a positive impression. Many praised individual physicians and nurses who had made a difference for them (although some, of course, also told of negative experiences). The same type of superlatives was hardly ever used about PAEHR.

Incomprehensible: Patients Are Not the Intended Reader of the Medical Records

Another reason why PAEHRs were not always perceived as attractive was that the language used in medical records could be hard to understand. Patient 7, for example, a retired cleaner, had never logged into Pasientjournal for the very reason that she did not think she would understand much of the contents:

> *"It has to be understandable, so they must not use a lot of foreign words that I don't know or Latin words for different things, this and that and that"* [Patient 7, a woman in her 60s; no PAEHR experience]

Patient 32 expressed a similar sentiment, even though she worked as an auxiliary nurse. When asked if she knew she could access her records, she answered:

> *"Yes, I do know that, but I am thinking: What's the point? The doctors' language ... I certainly do not understand much of that."* [Patient 32, a woman in her 50s; no PAEHR experience]

Many of the participants who had accessed their records also described how the medical records could be difficult to comprehend. Patient 10, for example, had accessed Pasientjournal to see what her doctor had written about the results of a sleep monitoring procedure she had undergone. She found the note, but she felt she could not really make much sense of it:

> *"You know, many of the documents in there, they are more like letters to doctors, or therapists, so there is a lot of inexplicable words, and I do not understand everything, I really don't."* [Patient 10, a woman in her 20s; PAEHR user]

XSL•FO
RenderX

Misery Oriented: Too Much Focus on Disease

Several study participants had reservations about PAEHRs because they wanted to avoid unnecessary focus on ill health. Patient 3, for example, a man in his 60s who had a chronic condition that demanded regular blood transfusions, explained it this way:

> *"I wouldn't say that I am totally uninterested in my disease; of course I'm not. But I'm not all that interested in it either, as long as everything is working as well as it does. So, I have chosen not to think about this disease too much. ... I think it is important for the total well-being, or the feeling of not digging oneself down in the disease and give it your main focus. I focus on all the things I can do, and that's a whole lot of things. ... I believe that it is important to be able to live with this kind of serious, chronic disease; that you do not go around thinking about it all the time. I believe that will make you feel like you are locked in inside yourself."* [Patient 3, a man in his 60s; no PAEHR experience]

Similarly, Patient 32, a woman in her 50s, also questioned what good might come from worrying about things she could not do anything about:

> *"You see, what good will it do to anyone around me if I were to dig down in my own misery? If this is something I will die from, then I want to live while I'm alive"* [Patient 32, a woman in her 50s; no PAEHR experience]

What these participants seemed to argue was that immersing oneself in disease and medical knowledge may not be a favorable way of dealing with illness and might even stand in the way of "good health."

Fear Provoking: Promoting Unnecessary Worry

Some study participants were concerned that reading the medical records would make them upset or worried. In their assessment, accessing their records had the potential to harm.

> *"No, if one starts to follow up on that stuff, one will start to speculate about all sorts of things, and then it certainly becomes worse. ... that is also why I never Google survival outcomes on stenting of heart vessels, that's for sure!"* [Patient 13, a man in his 50s; no PAEHR experience]

As Patient 15 pointed out,

> *"In some ways it feels good to be kind of ignorant, too. I do not want to read any reports before I know that this [the disease] has turned out good, in a way."* [Patient 15, a man in his 40s; no PAEHR experience]

Some participants who had used Pasientjournal explained how reading the health records had led to negative feelings. For example, a man suffering from incurable cancer explained how reading about his prognosis felt harder and more brutal than talking to his doctor about it:

> *"Even if it is information that I already know, it's sort of hard at times to sit down and read, when it's there, black on white, that they [the doctors] are not that optimistic anymore, you know. ... so it's kind of brutal."* [Patient 4, a man in his 50s; PAEHR user]

Avoiding the worries and negative feelings that they believed could result from being exposed to the illness focus in the health records constituted an important rationale for some of the participants for not engaging with the PAEHR.

Energy Demanding: Unsuitable When Ill Health Drains You of Your Vigor

When experiencing serious disease, patients may be indisposed of their normal capacities. Under such circumstances, some of the study participants had simply not had the energy to log onto Pasientjournal. In the case of Patient 5, this phase had lasted for several months after she was diagnosed with and went through surgery for breast cancer. Although she knew about the PAEHR and thought of it as an interesting service, she had not yet had the energy to use it 3 months after her surgery:

> *"I have read about that thing, that it is a possibility, but it has been so much information to digest and absorb that I did not reckon I had the time to log on there. ... have just been so overwhelmed with thoughts, with this disease, so I have just not had the energy"* [Patient 5, a woman in her 40s; no PAEHR experience]

Participants who had experienced serious disease talked about the overwhelming situation they suddenly had found themselves in when going from being healthy and resourceful to becoming a patient. Some also described how being a patient involved a lot of work. Patient 1, for instance, had just retired when she was diagnosed with cancer. She described how dealing with her new situation felt like a job with lots of tasks to take on:

> *"Everything was just so unfamiliar, and I did not know how to handle things when I became ill, and all the stuff that came along with the disease. ...This [being sick], it is actually a full-time job. With added extra hours!"* [Patient 1, a woman in her 60s; no PAEHR experience]

The "energy demanding" rationale against PAEHR use draws from the experience of illness as something that demands effort from the patient in many different ways. In such situations, the task of reading health records may be too much to deal with on top of everything else.

Cumbersome: Screen May Not Be Better Than Paper

PAEHR solutions make health records available to patients through the internet. While this is often touted as making information "easily available," some study participants felt that electronic access was a cumbersome thing. Some, especially in the older age groups, explained how navigating and engaging with new digital services could be an onerous task. Although many used personal computers and smartphones for other undertakings, they could find themselves struggling when they had to handle a new service on their own. Most patients had relatives who they turned to with "data troubles," but some felt uneasy bothering them too much. Patient 37, who regularly turned to her son for help with "technical stuff," explained:

XSL•FO
RenderX

> *"You know, my son, he is an expert on data, so he tries to explain it to me: 'Mother, you have to click here, and then this and that,' but then he has to run to catch the bus, and I'm like: 'What did he say???' Even before he's out the door, it's all gone for me."* [Patient 37, a woman in her 70s; no PAEHR experience]

Among the participants who questioned the "easiness" of digital solutions was Patient 12, a woman in her 70s. She had become upset when her pharmacy stopped giving out lists of prescriptions on paper and instead encouraged its customers to go online. She had logged on several times to check her prescriptions, but when the pharmacy did away with the new policy, she was relieved. Although she mastered the digital tool and could access her prescriptions, she found it much less complicated to be given a paper copy when she picked up her medicines. Patient 37 and Patient 12 are examples of patients who do not find that digital access represents progress and is "easier" than analog solutions.

A consequence of needing assistance to use a patient portal is that one must be willing to let the helper see the contents of one's medical records. Some study participants were not comfortable with this, and one explained that this was the main reason why she had never used the Pasientjournal.

Impoverishing: Resisting the Digital Transformation of Individual and Social Life

A final rationale against the use of PAEHRs was a general skepticism toward the ongoing digitalization of life worlds. Patient 7, for example—a woman in her 70s who was admitted to the hospital with newly discovered cancer—explained that she felt frustrated by how everyone was hooked on their screens and how this affected the social life in her family.

> *"I don't want to get into that stuff. I see everyone sitting there with their tablets all the time; they are completely absorbed. You lose the social part, I've said that from the beginning. And it's all...and it includes the kids as well, and the grandkids. They are all staring into a tablet, they all have a phone they are constantly playing on, or something. And I really resent much of that! Because you don't have the social engagement anymore."* [Patient 7, a woman in her 70s; no PAEHR experience]

Patient 7 had chosen not to engage in any digital affairs. She left that to her husband. They had a computer at home, but she never used it. "I'm not at all interested in that!" she stated, and she did not see the PAEHR as a service worth embracing. Patient 7 was the most prominent representative of an antidigital attitude in our study.

Other study participants were less vigorous in their rejection of digitalization but still expressed that they wanted to reduce the time spent in front of screens and computers, limiting such activities to strictly necessary chores. Patient 38, a man in his 70s, had heard of PAEHRs but had not yet used Pasientjournal. When asked if he would consider accessing his hospital health records, he was a bit hesitant, and said, "I guess I could do that." When pushed a bit further, he explained how he had experienced that computers interfered with his former work life as a shop owner:

> *"It took me away from the clients. Instead of being out in the shop and talking to customers, I was stuck at the back dealing with computer issues and troubles."* [Patient 38, a man in his 70s; no PAEHR experience]

Based on his experiences, Patient 38 felt a reluctance to engage too much with computers.

Discussion

Principal Findings

In this study into patients' rationales for not embracing online access to health records, we identified 8 main motivations for nonuse. These results expand the current knowledge of patients' perspectives on PAEHRs.

A fundamental reason why some study participants did not embrace PAEHRs in our study was that they did not think that reading their own medical record was particularly useful. It was unclear to them exactly what, if anything, reading medical notes would add to their understanding of their ailment, medical care, and/or condition. These views are in contrast to the more positive perspectives on the utility of PAEHRs reported in several previous studies [7,11] as well as by many of the participants with user experience in this study (to be published in another article). Participants who fronted the "unnecessary" rationale against PAEHRs felt that they got the necessary information from their doctors and did not seem to desire more medical information.

Several patients also preferred dialogic information given face to face rather than written information, which brings us to the second rationale identified: medical records are impersonal. Reading PAEHRs is an activity that does not typically involve human contact or interaction, and PAEHRs provide no 2-way interaction. As you read, there is no one to ask, no one to laugh or cry with, and no one to seek comfort and reassurance from. If considerations related to efficiency and costs should lead to human interaction being substituted by digital access to records that are not written primarily with the patient in mind, health care would surely risk becoming impoverished.

Many of our study participants (albeit not all) felt that their health care services were well-functioning and that their need for information and explanation was met in direct communication with their health care providers. One could hypothesize that PAEHRs might be perceived of as relatively more useful in a setting where patients were not as content with the services and the information they received. There may be some evidence to support this hypothesis. Surveys of PAEHR users in the United States have found that patients from underserved groups (less educated, non-White, older, and Hispanic patients, and individuals for whom English is a second language) are the ones most likely to report "major benefits" from reading medical notes [9].

An issue raised by study participants who had never used PAEHRs, as well as those who had used it either frequently or

XSL•FO
RenderX

infrequently, was that laypeople are not the intended audience for the documents accessible through PAEHRs. A survey of Norwegian PAEHR users previously found that understanding the content of the medical records is a challenge for many: 36% indicated that it was hard to understand what the documents were about, and 59% reported that they had difficulties understanding the medical terminology [7]. A Danish study that examined the language in 10 medical records with the aim of identifying "potential lay-friendliness, patient-centeredness, and patient empowerment" found that the records were written in highly specialized expert language dominated by expert terminology, expert syntax, expert presuppositions, and difficult abbreviations [19]. The authors concluded that "the majority of Danes will not fully understand their own e-records and will have a high potential for misunderstandings" [19]. In sum, few medical record entries are written with patients as the intended reader. Notes are typically prepared with other functions in mind and tend to be worded with the understanding that the medical file is a communication platform for professionals engaged in the care of the patient. The logic behind assessments, choices, and treatments is seldom explained explicitly; the text rests on the expectation that the reader is a health care professional with certain crucial preunderstandings. In short, the health record is not a resource dedicated to providing enlightenment, sympathy, or education to patients. Of course, this is not to say that the information in the records are never useful for patients, and many patients see PAEHRs as a good place to find information. Other studies have found that some patients use the PAEHR to remember what they talked about with their doctor or to prepare for an upcoming doctor's appointment [20]. However, as described in this study, the unfamiliar and professional atmosphere in the health record served as a barrier to PAEHR use for some patients, and they preferred using other sources for information about their health and illness. In the future, this barrier may be lowered by tools providing easy translation and explanations of medical terminology used in the PAEHR [21]. It must also be pointed out that Pasientjournal is currently only available within specialized health care services, and therefore the contents in the health records are often of a specialized and rather complex nature. One could hypothesize that the content of the records kept in general practice might be of a more accessible and understandable nature, and that the "incomprehensible" rationale might be less prominent in general practice settings than in hospitals.

Medical records may shift one's focus from healthy and wholesome parts of life to sickness and worries about ill health. Avoiding unnecessary focus on sickness and pathology seemed to be an important strategy for some of our study participants. Similar findings were described in a qualitative study of self-management practices among patients with chronic conditions. Van de Bovenkamp and Dwarswaard [22] explained how "self-management is very much shaped by patients' ideas of the good life." A study by Henwood et al [23] of health information–seeking practices among women in midlife similarly identified resistance against the idea that patients should seek out medical information by themselves rather than trusting their doctors to equip them with it. The quotation used by the authors in the article's title—"ignorance is bliss sometimes"—resonates with some of the patients' views in this study: more knowledge is not necessarily for the better. Moreover, some of our study participants clearly also actively resisted the biomedical version of their conditions as presented in the medical records [24]. To them, illness was more than the biomedically relevant information written down in the PAEHR. Illness included feelings, everyday challenges, and encouragements—in short, components of a life that the biomedically dominated health records were lacking.

In addition, some participants stayed away from PAEHRs because they felt that medical records might have the capacity to induce fear and anxiety, and some described situations in which reading the medical records had led to unwanted emotional reactions. This is reflected in a Norwegian survey of PAEHR users as well: 20% of respondents indicated that they had become worried by reading health records [7]. Clearly, how much and what kind of information one seeks at what time will depend on individual factors and previous experience and may also change over time. Some PAEHR nonusers preferred to stay somewhat "ignorant" and rely on the doctor to curate the necessary information for them.

The "new patient role" has been critiqued for not recognizing that patients oftentimes are unable and weakened—temporarily or permanently—when they are sick, and that this affects their ability to act as they otherwise would and make "rational" choices [25,26]. Reading and making sense of the health records is a complicated task (at times even for health professionals). It requires concentration over time and cognitive capacity to absorb information. Lack of concentration can be a temporary or enduring result of a patient's condition or a side effect of medication—all described by participants in this study. For some of the interviewees, this was exactly why they had not accessed their health records. Moreover, as also described by participants in this study, becoming a patient may resemble taking on a new full-time job. As Mol argued in *The Logic of Care* [26], the notion of the neoliberal "patient citizen" (or "active patient" in the PAEHR context) presupposes a nonsuffering person who can make well-argued individual choices, whereas in reality the patient is often weakened by disease and emotional distress that delimits his or her ordinary capacities. The experiences of our participants related to Mol's argument that good care implies engagement in ongoing attempts to adapt knowledge and technologies in a collaboration involving patients and their health care providers as well as their diseased bodies and complex lives.

Certainly, technological obstacles also played a role in our study. To make use of PAEHRs, some of our participants felt they would have to ask for help from relatives or friends, thereby revealing the dependency situation digital health records may impose on patients. Wyatt et al [27] have described how using the internet for health information seeking may require access to competent helpers that can guide less-experienced users in finding and making sense of the information. This was also a concern for some of the participants in this study, and perhaps not surprisingly, especially among some of those who were older. Several participants used the internet for other purposes but were not confident they could handle a new service or task without assistance. Many were hesitant to ask for help from relatives or friends who were more digitally competent ("warm

XSL•FO
RenderX

experts," to use Bakardjieva's term [28]). Depending on the nature of a person's current and former health-related issues, the record can contain information that patients feel is sensitive and too private to share with others, even (and sometimes especially) with close relatives. The threshold to ask for help from "warm experts" might therefore be higher in the case of PAEHRs than for many other digital tasks.

Finally, some of the study participants were generally critical toward the digitalization of individual and social life. In a world characterized by optimism on behalf of digital innovations, this kind of general resistance is at times dismissed as reactionary, and its representatives may at times seem to be rendered invisible in public discourse. However, our study revealed sound and rational arguments for staying outside of digital life. In health care services, this group must also be taken into account to ensure that all patients are equally taken care of within the health care system.

Limitations and Further Research

This article builds on qualitative research and we are therefore in no position to know how common the rationales against PAEHR use identified by our study participants may be. Moreover, our study was limited in time, geography, and size, and further work in the same location or elsewhere could bring out additional rationales of relevance and importance. Furthermore, there is a potential for researcher bias here as we know that many health personnel are skeptical toward PEAHRs and the 3 authors are all physicians. However, all 3 of us have chosen other careers than the clinical as our main jobs and do not primarily identify as clinicians. None of us had preconceived ideas of a potential negative impact before embarking on this study.

There is a need for more research on rationales for not using PAEHRs. Surveys quantifying different rationales for and against PAEHR use, as well as studies of the demography of users and nonusers, are important. There is also a need for more qualitative studies exploring potential differences between specialist health care settings and primary health care settings, as well as possible factors contributing to positive outcomes from and experiences with PAEHRs.

Practical Implications

This study demonstrates that patients may have good reasons for not using PAEHRs and could help guide how PAEHRs are promoted among patients. While PAEHRs are a beneficial service for many patients, it may not be the case for all patients. Health care providers should take this into consideration and should avoid assuming that information is received or understood because it is available via PAEHRs. Moreover, when PAEHRs are promoted, one should avoid making the assumption that online access to health records will be useful for all or should be used by all.

Conclusions

It is often assumed that barriers against PAEHR use are mostly practical (such as lack of access to hardware, slow internet connections, etc) and that once such hurdles are removed, PAEHRs will be widely embraced. In this paper, however, we have demonstrated that patients may have many good reasons not to celebrate or adopt this kind of service. Assessed as a potential source of insight and understanding, some find that PAEHRs are less than perfect because they are devoid of human interaction and contact, use language that they do not understand, promote a 1-sided focus on misery, may lead to unnecessary fear, demand more energy than one may have when sick, threaten to expose one's secrets, and can be cumbersome to make work. Perhaps more than anything else, what the study participants highlighted was that the information in health records is not in a format that is genuinely written for patients. To the contrary, health records are primarily a peephole into biomedical perspectives on the patient's misfortune and ill health. What one gets access to through PAEHRs is often biomedical explanatory models written in biomedical language for biomedically trained professionals. This study demonstrates that not all patients view this as either helpful or favorable, and promoting the view that PAEHRs are something all patients ought to engage with might lead some patients to feel pushed into a domain they are not comfortable in.

Acknowledgments

Fieldwork, transcription, and coding were conducted by the first author (HSV). All authors contributed to the design of the study, analysis of the data, and writing and editing of the manuscript. The study was funded by the University of Oslo. We would like to thank the hospital that kindly invited us to conduct the research and to express our deep gratitude to all of the study participants who contributed to this study with their time, thoughts, experiences, and perspectives.

Conflicts of Interest

None declared.

References

1. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient Access to Electronic Health Records: Differences Across Ten Countries. Health Policy Techn 2018 Jun;7(2):224-256 [FREE Full text] [doi: 10.1016/j.hlpt.2017.11.003]
2. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]

XSL•FO
RenderX

3. Vermeir P, Degroote S, Vandijck D, Van Tiggelen H, Peleman R, Verhaeghe R, et al. The patient perspective on the effects of medical record accessibility: a systematic review. Acta Clin Belg 2017 Jun;72(3):186-194. [doi: 10.1080/17843286.2016.1275375] [Medline: 28056665]
4. Mold F, de Lusignan S. Patients' Online Access to Their Primary Care Electronic Health Records and Linked Online Services: Implications for Research and Practice. J Pers Med 2015 Dec 04;5(4):452-469 [FREE Full text] [doi: 10.3390/jpm5040452] [Medline: 26690225]
5. Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A. "Nothing About Me Without Me": An Interpretative Review of Patient Accessible Electronic Health Records. J Med Internet Res 2015 Jun 29;17(6):e161 [FREE Full text] [doi: 10.2196/jmir.4446] [Medline: 26123476]
6. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient Portals Facilitating Engagement With Inpatient Electronic Medical Records: A Systematic Review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
7. 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]
8. 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]
9. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
10. 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]
11. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J 2018 Jun;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]
12. Use of ICT in households. Statistisk Sentralbyrå. 2020 Sep 16. URL: https://www.ssb.no/ikthus [accessed 2021-05-04]
13. OECD. Digital Government Review of Norway: Boosting the Digital Transformation of the Public Sector. Paris: OECD Publishing; 2017.
14. Meld.St. nr 9 (2012-2013). Én innbygger - én journal. Oslo: Helse- og omsorgsdepartementet; 2012.
15. Sørensen T, Johansen MA. Developing and Implementing Patients' Full-Scale Electronic Access to Their Health Record. Stud Health Technol Inform 2016;228:85-89. [Medline: 27577347]
16. Helse Nord RHF. Åse viser deg hvordan du bruker sykehusjournalen på nett. URL: https://www.youtube.com/watch?v=27waHHrWeR0 [accessed 2020-06-30]
17. Brinkmann S, Kvale S. InterViews: Learning the Craft of Qualitative Research Interviewing. 3rd ed. Thousand Oaks, Calif: SAGE Publications, Inc; 2014.
18. Tjora A. Kvalitative forskningsmetoder i praksis. 3. utg. Oslo: Gyldendal akademisk; 2017.
19. Hansen M, Zethsen K. Is Electronic Access to Medical Records an Empowering and Patient-centered Initiative? – a Qualitative Contextual and Linguistic Analysis of Danish Electronic Records. Hermes 2018 Dec 22(58):157-173 [FREE Full text] [doi: 10.7146/hjlcb.v0i58.111683]
20. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
21. Lalor JP, Woolf B, Yu H. Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers. J Med Internet Res 2019 Jan 16;21(1):e10793 [FREE Full text] [doi: 10.2196/10793] [Medline: 30664453]
22. van de Bovenkamp HM, Dwarswaard J. The complexity of shaping self-management in daily practice. Health Expect 2017 Oct;20(5):952-960 [FREE Full text] [doi: 10.1111/hex.12536] [Medline: 28152248]
23. Henwood F, Wyatt S, Hart A, Smith J. 'Ignorance is bliss sometimes': constraints on the emergence of the 'informed patient' in the changing landscapes of health information. Sociol Health Illn 2003 Sep;25(6):589-607 [FREE Full text] [doi: 10.1111/1467-9566.00360] [Medline: 12919447]
24. Kleinman A. The Illness Narratives: Suffering, Healing, and the Human Condition. New York: Basic Books; 1988.
25. Lupton D. Consumerism, reflexivity and the medical encounter. Soc Sci Med 1997 Aug;45(3):373-381. [doi: 10.1016/s0277-9536(96)00353-x] [Medline: 9232732]
26. Mol A. The Logic of Care: Health and the Problem of Patient Choice. London: Routledge; 2008:9780415453424.
27. Wyatt S, Henwood F, Hart A, Smith J. The digital divide, health information and everyday life. New Media Soc 2005 Apr 01;7(2):199-218 [FREE Full text] [doi: 10.1177/1461444805050747]
28. Bakardjieva M. Becoming a Domestic Internet User. In: Internet Society: The Internet in Everyday Life. London: Sage Publications Ltd; 2005:92-116.

XSL•FO
RenderX

Abbreviations

PAEHR: patient-accessible electronic health record

Edited by G Eysenbach; submitted 04.09.20; peer-reviewed by M Hägglund, B Lo, J Pecina; comments to author 30.10.20; revised version received 23.02.21; accepted 03.04.21; published 28.05.21.

Please cite as:
Valeur HS, Lie AK, Moen K
Patient Rationales Against the Use of Patient-Accessible Electronic Health Records: Qualitative Study
J Med Internet Res 2021;23(5):e24090
URL: https://www.jmir.org/2021/5/e24090
doi:10.2196/24090
PMID:34047711

©Hanne Støre Valeur, Anne Kveim Lie, Kåre Moen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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

Testing a Novel Web-Based Neurocognitive Battery in the General Community: Validation and Usability Study

Riley Capizzi[1], BSc; Melissa Fisher[2], PhD; Bruno Biagianti[3], MD; Neelufaer Ghiasi[2], BA; Ariel Currie[2], MBS; Karrie Fitzpatrick[2], PhD; Nicholas Albertini[2], BSc; Sophia Vinogradov[2], MD

[1]Department of Psychology, Temple University, Philadelphia, PA, United States

[2]Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, United States

[3]Department of Research and Development, Posit Science Corporation, San Francisco, CA, United States

Corresponding Author:
Riley Capizzi, BSc
Department of Psychology
Temple University
1701 N 13th St
Philadelphia, PA, 19122
United States
Phone: 1 (847) 997 4148
Email: capizzi@temple.edu

Abstract

Background: In recent years, there has been increased interest in the development of remote psychological assessments. These platforms increase accessibility and allow clinicians to monitor important health metrics, thereby informing patient-centered treatment.

Objective: In this study, we report the properties and usability of a new web-based neurocognitive assessment battery and present a normative data set for future use.

Methods: A total of 781 participants completed a portion of 8 tasks that captured performance in auditory processing, visual-spatial working memory, visual-spatial learning, cognitive flexibility, and emotional processing. A subset of individuals (n=195) completed a 5-question survey measuring the acceptability of the tasks.

Results: Between 252 and 426 participants completed each task. Younger individuals outperformed their older counterparts in 6 of the 8 tasks. Therefore, central tendency data metrics were presented using 7 different age bins. The broad majority of participants found the tasks interesting and enjoyable and endorsed some interest in playing them at home. Only 1 of 195 individuals endorsed *not at all* for the statement, "I understood the instructions." Older individuals were less likely to understand the instructions; however, 72% (49/68) of individuals over the age of 60 years still felt that they *mostly* or *very much* understood the instructions.

Conclusions: Overall, the tasks were found to be widely acceptable to the participants. The use of web-based neurocognitive tasks such as these may increase the ability to deploy precise data-informed interventions to a wider population.

(J Med Internet Res 2021;23(5):e25082) doi:10.2196/25082

KEYWORDS

cognition; normative; remote; digital; online; web-based; BrainHQ; Posit Science Corporation

Introduction

Background

For decades, neuropsychological methods have been leveraged to understand and characterize the relative strengths and weaknesses of individuals experiencing an array of neuropsychiatric syndromes [1-5]. These profiles have also been shown to predict future deterioration in Alzheimer disease and other cognitive and functional outcomes in schizophrenia, bipolar disorder, depression, and traumatic brain injury [6-8]. However, traditional paper and pencil assessments are lengthy and expensive and require considerable training on the part of the assessor. Given these concerns and the rising availability of personal mobile technologies, a movement to capture reliable cognitive functioning digitally has begun [9-12]. Although

XSL•FO
RenderX

remote cognitive technologies are still budding, they promise several benefits in clinical and research settings.

Remote platforms greatly increase the accessibility of psychological testing. The National Institute of Mental Health has established a priority to reach typically underserved populations, such as individuals living with limited physical mobility or in rural areas [13]. Digital measures allow testing to be conducted within the comfort of one's own home, thereby increasing the ability to serve historically difficult-to-reach patients. Remote testing would also likely reduce the time providers spend scheduling and carrying out in-person assessments. These benefits may also help to reduce assessment costs. Accessibility concerns have taken center stage during the current global pandemic, highlighting the need for reliable, easy-to-use, and remote measures.

In addition to opportunities for accessibility, digital remote methods for cognitive testing encourage the use of longitudinal assessments. Repeated measurements provide a wealth of information above and beyond individual snapshots of performance [14]. A study of individuals at clinical high risk for conversion into a psychotic disorder found that although baseline measures of cognition helped differentiate at-risk individuals from healthy participants, cognitive trajectories followed for 2 years differentiated converters from nonconverters [15]. Another study of healthy older adults found that repeated memory assessments pushed to a handheld device were widely accepted and showed associations with the hippocampal brain structure, whereas typical baseline measures of cognition did not [16].

Although in-person neuropsychological methods unquestionably have their benefits and function optimally in distraction-free testing environments, the generalizability of data collected in these unique settings has been questioned [17]. There is a growing body of literature evaluating assessments that can be administered in people's daily lives, thus potentially capturing more meaningful data on how they function within their own environments [18,19]. For example, a clinician may be more interested in how their patient performs cognitively while completing their night shift at work rather than in a highly controlled testing center at midday.

Digital methods also provide a means to measure important metrics such as reaction time (RT) and trial-by-trial performance, which have also shown associations with clinical outcomes [20-22]. These potential benefits, alongside the added ability of electronic health record integration, may further facilitate person-centered practice and move us closer toward the goal of precision medicine [23]. However, little to no normative data currently exists for these task batteries, limiting the interpretability of new findings [24].

A particular novel mobile assessment platform implemented by the Posit Science Corporation has shown recent utility capturing cognitive performance differences common to psychiatric disorders, such as schizophrenia and depression, using a neuroscience-informed approach to task development [25,26]. These tasks were designed to capture the variability associated with aberrant neural mechanisms underlying an individual's ability to flexibly acquire new information and adapt to changing cognitive and emotional demands, integral processes for normative cognitive functioning and general learning ability [26-29]. Sound sweeps, a task of auditory processing speed, showed improvements during a cognitive training intervention in individuals with recent-onset and chronic schizophrenia [30]. The greatest improvement was observed during the first 20 hours of training, followed by a plateau at subsequent assessment points. Intraindividual variability in the time taken to reach this plateau was also associated with the likelihood that the intervention would generalize to other untrained cognitive domains, suggesting a potential target-mediated treatment response. Another study found that improvements in sound sweeps performance after cognitive training were correlated with gains in working memory and global cognition [31]. Despite the apparent utility of these tasks, they have yet to be adequately evaluated in normative data samples.

Objectives

Therefore, we set out to test a battery of 8 web-based cognitive tasks developed by Posit Science in samples drawn from the general community. Sensory perception, social cognition, and executive functions were digitally assessed in-person with state fair attendees and remotely through testing of college students. In addition, the usability and feasibility of this new battery were investigated in a subset of participants.

Methods

Participants

All study procedures were approved by the institutional review board of the University of Minnesota (UMN). Recruitment and study participation took place either in-person at the Minnesota State Fair (MSF) or remotely for course credit at UMN. The MSF is the second-most highly attended state fair in the United States, with approximately 2 million Midwesterners visiting every year [32]. State fairgoers represent a wide array of demographic backgrounds [33]. MSF participants were asked to participate if they were aged between 18 years and 80 years (inclusive), and UMN students were asked to participate if they were aged between 18 years and 40 years (inclusive). Inclusion criteria for all participants were (1) no visual, auditory, or motor impairments that would prevent completion of the assessments; (2) fluent and literate in English; and (3) no use of illicit substances or alcohol over the prior 8 hours. We received cognitive data from 816 participants and usability survey responses from 219 participants who participated in a substudy. However, cognitive data from 35 participants who took the battery multiple times were omitted from the primary cognitive analyses below, and survey data were omitted for another 24 participants who completed usability questionnaires multiple times. Therefore, 781 participants were included in the primary analyses, and survey data from 195 participants were inspected (Table 1).

XSL•FO
RenderX

Table 1. Participant demographics by sample.

Demographics	Sample		Total (N=781)
	College (n=202)	State fair (n=579)	
Gender[a]			
Female, n (%)	144 (71.6)	338 (58.6)	482 (62.0)
Male, n (%)	57 (28.4)	237 (41.1)	294 (37.8)
Intersex, n (%)	0 (0.0)	2 (0.3)	2 (0.3)
Age, mean (SD)	20.12 (2.28)	44.82 (17.72)	38.4 (18.7)
Years of education, mean (SD)	14.59 (1.5)	16.57 (2.77)	16.1 (2.65)
Grade point average, mean (SD)[b]	3.45 (0.35)	N/A[c]	3.45 (0.35)
Occupation level, mean (SD)[d]	5.16 (1.68)	3.06 (1.84)	3.81 (2.05)

[a]Three participants preferred not to respond.

[b]Grade point average only collected in the undergraduate samples.

[c]N/A: not applicable.

[d]Hollingshead two-factor index: occupational scale.

Procedures

After receiving informed consent, the study participants provided demographic information and completed a battery of cognitive tasks. College students participated remotely via their own personal devices (eg, tablets or personal computers) and were asked to find a stable internet connection to complete the tasks in a quiet, private environment using headphones, preferably over-the-ear headphones. MSF participants completed the study procedures in an enclosed fair structure using lab iPads (Apple Inc) and over-the-ear headphones. Participants were randomly assigned to complete 3 to 4 of the 8 cognitive assessments. State fair participants completed a subset of the measures because of the time limit set by the Driven to Discover State Fair program. A maximum of 30 minutes was allowed for giving consent and the completion of demographic information, cognitive measures, and questionnaires. Participants were randomly assigned to complete 1 of 2 cognitive batteries that were equivalent in length and comparable in terms of the number of auditory, visual, and social cognition tests.

Online Neurocognitive Assessments

Our work aims to translate measures from basic cognitive neuroscience into short, computerized assessments of discrete cognitive processes that individuals can easily complete with minimal assistance in various settings [25]. These assessments are designed to enable the interpretation of specific deficits that could signal that an individual is experiencing cognitive difficulties and impaired learning ability.

The first step in the development of the assessment suite was to decide on the cognitive domains and processes that are known to play a critical role in an individual's ability to learn new information, to interact adaptively with cognitive and emotional challenges in the environment, and to adapt to new learning demands. In line with the principles of team science, we integrated theoretical perspectives, technical expertise, and empirical knowledge drawn from a team of cognitive neuroscientists working in human and animal model systems, clinical researchers, and preclinical translational behavioral neuroscientists. Through a consensus-building process, we identified 3 critical neural domains: (1) perceptual processing (sound sweeps and beep seeker), (2) executive functioning (bubble pop, pathfinder, and mind bender), and (3) social-emotional processing (face to face, tap the emotion, and emotional face). Within each domain, we identified constructs, that is, a definable cognitive process that could be measured at the behavioral level and for which there existed clearly hypothesized and measurable neural-circuit mechanisms (eg, for executive functioning, set-shifting). We identified a cognitive neuroscience paradigm that could selectively and parametrically measure each of these constructs at the behavioral level. Guided by item response theory, these assessments use adaptive testing models to adjust the difficulty level according to the user, thereby reducing the duration of the test [34]. Online neurocognitive assessments (ONAs) take an average of 184 (SD 201) seconds to complete. Instructions on sampling the ONAs are presented in Multimedia Appendix 1.

Beep Seeker: Auditory Discrimination and Sensory Memory

Beep seeker is an auditory discrimination task in which participants are presented with a target tone and are asked to identify it in later trials amidst 2 other distractor tones. Accurate identification of the target tone prompts subsequent trials with more similar distractor tones. A linear staircase method was used to identify the participant's discrimination threshold on a scale of 1 to 15, with higher scores indicating better performance.

Sound Sweeps: Auditory Perception and Processing Speed

Sound sweeps is an auditory perception task in which participants are presented with 2 consecutive tones that may either sweep from a low to high pitch or high to low pitch. The participants were asked to identify the direction of each sweep. The sweep sounds' speed varied according to trial performance,

XSL•FO
RenderX

and participant performance was measured as $\log_{10}$(average RT in seconds).

Bubble Pop: Visual-Spatial Working Memory

Bubble pop is a working memory task in which participants are asked to follow a set of target bubbles that independently move around the screen alongside other distractor bubbles. The number of target bubbles varied with the performance. Accuracy was measured as the number of bubbles correctly tracked using a 2-up, 1-down staircase method.

Pathfinder: Visual-Spatial Learning

Pathfinder is a learning test. A path with 15 nodes is presented. The participant then attempted to recreate the path from memory. If a node is missed, the trial ends, and the path is shown again (5 trials in total). Accuracy was measured as the percentage of nodes correctly recalled out of the total number of nodes.

Mind Bender: Cognitive Flexibility

Mind bender is a task that instructs participants to identify images that follow an established rule around other pictures that violate the rule. Performance was measured as $\log_{10}$(correct trial RT in milliseconds).

Tap the Emotion: Emotion Detection and Inhibitory Control

Tap the emotion requires the participant to tap the screen when a happy or sad face appears and inhibits the prepotent response to tap when presented with a neutral face. Performance was measured as the mean accuracy of neutral trials.

Face to Face: Emotion Identification

A face is shown with a specific emotion, followed by a series of faces with various emotions. Participants were asked to identify the target emotion first shown among the series of faces. The task varies in difficulty by changing the number of emotions presented and the number of faces to choose from. Performance was measured as $\log_{10}$(duration of target stimulus presentation).

Emotional Face: Inhibitory Control of Emotion

Emotional face provides a measure of executive attention, in which an expressive face is shown and overlaid with a congruent or incongruent word (Stroop effect). The increased RT to incongruent stimulus combinations captures the capacity for conflict resolution. Performance was measured by subtracting the mean RT of the correct congruent trials from the mean RT of the correct incongruent trials.

Usability Questionnaire

This lab-designed measure asked participants to indicate how much they agreed with a series of 5 statements using a 5-point Likert scale ranging from *not at all* to *very much*. The specific questions are shown in Multimedia Appendix 2.

Statistical Analysis

All main outcomes were screened for normality. Although some measures showed skewness and kurtosis (eg, beep seeker, tap the emotion, and face to face), all had adequate variance. The task distributions are shown in Multimedia Appendix 1. Nonparametric statistics were subsequently used for beep seeker, tap the emotion, and face to face. Outliers greater than 2.5 SD from the mean were winsorized and represented less than 1% of the data. For ease of comparison, outputs from sound sweeps, mind bender, and face to face were transformed so that higher scores indicated better performance.

General tendencies, dispersion metrics, and associations with demographic variables were calculated for each ONA. A total of 2 individuals identified as intersex were excluded from the gender comparison of ONA performance because of insufficient power. Given that 6 of 8 ONAs showed significant associations with age, task performance was also summarized across 7 different age ranges. To investigate concurrent and discriminant validity across ONAs, a Spearman rank-order correlation matrix was summarized; however, a factor analysis was not conducted because of inefficient overlap of measures completed across samples and a subsequent lack of power.

Results

In total, 781 individuals completed at least one ONA (college, n=202; MSF, n=579). Table 2 presents the means, SD, and ranges for each task.

XSL•FO
RenderX

Table 2. Normative statistics by online neurocognitive assessment.

Assessments	Participant, n (%)	Task metric, mean (SD)	Minimum[a]	Maximum[b]
Beep seeker	269 (34.4)	5.01 (4.63)	1.00	15.00
Sound sweeps[c]	269 (34.4)	0.98 (0.33)	0.00	1.70
Bubble pop	293 (37.5)	5.07 (1.20)	1.00	7.80
Path finder	323 (41.4)	52.32 (24.97)	1.00	100.00
Mind bender[c]	256 (32.8)	−9.09 (1.03)	−11.60	−6.80
Tap the emotion	426 (54.5)	78.48 (17.88)	20.00	100.00
Face to face[c]	339 (43.4)	−2.03 (0.42)	−3.45	−1.50
Emotional face	252 (32.3)	53.19 (80.11)	−169.00	342.90

[a]Minimum value refers to the lowest task metric across participants on a given task.

[b]Maximum value refers to the highest task metric across participants on a given task.

[c]Raw task output was inverted so that higher scores indicated better performance.

Over 96.4% (188/195) of participants found the ONAs to be at least *a little bit* interesting enjoyable. Almost 91.8% (179/195) of individuals found the tasks at least *a little bit* easy. Only 1 participant endorsed *not at all* to the statement, "I understood the instructions." The vast majority of participants found the ONA instructions easy to understand at the level *of a little bit* or more. Age was found to be negatively associated with the extent to which people agreed that they understood the instructions (ρ=−0.21; *P*=.004). Still, over 72% (49/68) of participants over the age of 60 years reported that they *mostly* or *very much* aligned with the statement, "I understood the instructions." In addition, when asked how much people agreed with the statement, "I would play these games at home," 69.2% (135/195) endorsed *a little bit* or more.

Task relationships with demographic variables such as gender, age, and education are presented in Table 3 and those with occupation level and grade point average (GPA) are presented in Table 4. Male participants outperformed female participants on beep seeker (Wilcoxon rank-sum [W]=6927; *P*=.01), sound sweeps (two-tailed *t* test: $t_{163.65}$=−2.12; *P*=.04), and path finder (two-tailed *t* test: $t_{264.93}$=−4.11; *P*<.001). Age was negatively correlated with task performance on sound sweeps (Pearson correlation [*r*]=−0.33; *P*<.001), bubble pop (*r*=−0.50; *P*<.001), path finder (*r*=−0.38; *P*<.001), mind bender (*r*=−0.41; *P*<.001), tap emotion (ρ=−0.24; *P*<.001), and face to face (*r*=−0.19; *P*<.001). Given the consistent association with age, task performance is displayed across 7 age bins, as shown in Tables 5 and 6. Counterintuitively, fewer years of education were associated with better performance on bubble pop (*r*=−0.13; *P*=.03), mind bender (*r*=−0.13; *P*=.04), and tap the emotion (ρ=−0.16; *P*=.001); however, when controlling for age, these relationships were not significant. Higher levels of occupation were significantly associated with an elevation in beep seeker (ρ=0.15; *P*=.02). Lower levels were associated with better performance on sound sweeps (*r*=−0.16; *P*=.03), bubble pop (*r*=−0.17; *P*=.01), path finder (*r*=−0.24; *P*<.001), mind bender (*r*=−0.21; *P*=.001), and tap the emotion (ρ=−0.20; *P*<.001), but when age was accounted for, only bubble pop remained associated ($F_{1,227}$=4.12; *P*=.04). GPA was not significantly correlated with any of the ONAs.

XSL•FO
RenderX

Table 3. Association of online neurocognitive assessments with demographic variables (gender, age, and education).

Assessments	Demographics											
	Gender				Age (years)				Education (years)			
	Student two-tailed *t* test		Wilcoxon ranked sum		Pearson correlation		Spearman rank-order correlation		Pearson correlation		Spearman rank-order correlation	
	t test (*df*)	*P* value	W	*P* value	*r*	*P* value	ρ	*P* value	*r*	*P* value	ρ	*P* value
Beep seeker	__[a]	—	6927	.01	—	—	0.10	.11	—	—	0.10	.11
Sound sweeps[b]	−2.12 (163.65)	.04	—	—	−0.33	<.001	—	—	<0.001	.99	—	—
Bubble pop	−1.09 (185.61)	.28	—	—	−0.50	<.001	—	—	−0.13	.03	—	—
Path finder	−4.11 (264.93)	<.001	—	—	−0.38	<.001	—	—	−0.10	.07	—	—
Mind bender[b]	−1.39 (145.12)	.17	—	—	−0.41	<.001	—	—	−0.13	.04	—	—
Tap the emotion	—	—	19,480	.28	—	—	−0.24	<.001	—	—	−0.16	.001
Face to face[b]	—	—	12,984	.85	—	—	−0.19	<.001	—	—	−0.01	.85
Emotional face	1.70 (223.3)	.09	—	—	−0.03	.61	—	—	0.01	.82	—	—

[a]Nonparametric test provided given violation of normality.

[b]Raw task output was inverted so a larger output indicated better performance.

Table 4. Association of online neurocognitive assessments with demographic variables (occupation level and grade point average).

Assessments	Demographics							
	Occupation level				Grade point average			
	Pearson correlation		Spearman rank-order correlation		Pearson correlation		Spearman rank-order correlation	
	r	*P* value	ρ	*P* value	*r*	*P* value	ρ	*P* value
Beep seeker	__[a]	—	0.15	.02	—	—	0.07	.57
Sound sweeps[b]	−0.16	.03	—	—	0.08	.39	—	—
Bubble pop	−0.17	.01	—	—	0.04	.67	—	—
Path finder	−0.24	<.001	—	—	0.13	.26	—	—
Mind bender[b]	−0.21	.001	—	—	0.08	.41	—	—
Tap the emotion	—	—	−0.20	<.001	—	—	0.02	.83
Face to face[b]	—	—	−0.09	.10	—	—	0.01	.93
Emotional face	−0.04	.56	—	—	0.04	.74	—	—

[a]Nonparametric test provided given violation of normality.

[b]Raw task output was inverted so a larger output indicated better performance.

XSL•FO
RenderX

Table 5. Normative data by age bin for ages 17-50 years.

Assessments	Age bin (years)							
	17-20		21-30		31-40		41-50	
	Participant, n	Mean (SD)	Participant, n	Mean (SD)	Participant, n	Mean (SD)	Participant, n	Mean (SD)
Beep seeker	76	3.76 (4.06)	63	5.11 (4.57)	22	6.91 (5.87)	32	6.50 (4.94)
Sound sweeps	102	1.08 (0.29)	62	1.04 (0.32)	19	0.84 (0.33)	22	0.88 (0.27)
Bubble pop	94	5.59 (0.95)	68	5.41 (1.13)	23	5.04 (1.01)	33	4.93 (1.17)
Path finder	91	61.86 (21.22)	70	59.79 (25.02)	25	55.16 (24.87)	33	46.18 (21.99)
Mind bender	97	−8.7 (0.95)	56	−8.9 (0.94)	20	−8.82 (0.96)	23	−9.59 (0.84)
Tap the emotion	136	86.29 (10.91)	108	74.76 (20.27)	32	72.56 (21.05)	39	74.32 (18.31)
Face to face	154	−1.96 (0.41)	80	−1.95 (0.38)	22	−2.17 (0.35)	25	−2.09 (0.42)
Emotional face	62	56.42 (76.89)	63	53.82 (73.4)	24	67.65 (91.75)	32	36.49 (64.09)

Table 6. Normative data by age bin for ages 51-80 years.

Assessments	Age bin (years)					
	51-60		61-70		71-80	
	Participant, n	Mean (SD)	Participant, n	Mean (SD)	Participant, n	Mean (SD)
Beep seeker	35	6.29 (4.93)	34	3.97 (3.66)	7	3.71 (4.11)
Sound sweeps	28	0.89 (0.28)	29	0.79 (0.38)	7	0.7 (0.40)
Bubble pop	32	4.57 (1.08)	33	3.93 (0.99)	10	3.68 (0.92)
Path finder	49	45.96 (25.25)	45	34.98 (21.28)	10	35.7 (21.89)
Mind bender	28	−10.05 (1.07)	28	−9.62 (0.74)	4	−9.43 (0.48)
Tap the emotion	51	78.80 (14.93)	48	75.6 (18.52)	12	62.92 (25.71)
Face to face	30	−2.11 (0.39)	24	−2.33 (0.56)	4	−2.29 (0.25)
Emotional face	35	55.09 (73.18)	30	52.26 (107.65)	6	38.07 (110.4)

Most visual tasks were significantly associated with one another, with Spearman rho coefficients ranging from 0.19 to 0.42, indicating some expected shared variance and evidence for the measurement of unique constructs (Table 7). Beep seeker and bubble pop, which measure similar memory processes across auditory and visual domains, were also correlated with one another; however, the Spearman rho was only 0.28, suggesting that they largely capture distinct constructs. In addition, although significant, sound sweeps was found to be only minimally associated (ρ=0.16-0.36) with other theoretically distinct constructs (eg, visual-spatial learning and emotion detection).

XSL•FO
RenderX

Table 7. Online neurocognitive assessments' correlation matrix.

Assessments	Beep seeker	Sound sweeps	Bubble pop	Path finder	Mind bender	Tap the emotion	Face to face
Sound sweeps							
ρ[a]	N/A[b]	—[c]	—	—	—	—	—
P value	N/A	—	—	—	—	—	—
Bubble pop							
ρ	0.28	0.18	—	—	—	—	—
P value	.002	.05	—	—	—	—	—
Path finder							
ρ	0.1	0.36	N/A	—	—	—	—
P value	.22	<.001	N/A	—	—	—	—
Mind bender							
ρ	0.24	0.31	0.19	0.23	—	—	—
P value	.05	<.001	.05	.005	—	—	—
Tap the emotion							
ρ	0.02	0.16	0.22	0.25	0.42	—	—
P value	.83	.04	.001	.006	<.001	—	—
Face to face							
ρ	0.09	0.20	0.14	0.23	0.23	0.19	—
P value	.43	.006	.15	.003	.001	.005	—
Emotional face							
ρ	−0.21	N/A	−0.09	0.11	N/A	0.07	−0.05
P value	.005	N/A	.35	.38	N/A	.32	.55

[a]ρ: Spearman's rank correlation coefficient.

[b]N/A: not applicable; the tasks were not included in the same battery.

[c]Redundant information.

Discussion

Principal Findings and Significance

In this study, we tested a new web-based neurocognitive assessment battery for individuals from the general community. Normative metrics were collected across 8 tasks measuring auditory discrimination, auditory perception, visual-spatial working memory, visual-spatial learning, cognitive flexibility, emotion detection, emotion identification, and inhibitory control of emotion.

Participants found the assessments to be widely acceptable across a range of ages. Age was negatively associated with the extent to which participants were able to understand task instructions; however, the broad majority of individuals in older age groups found the tasks understandable. Still, future iterations of these tasks should consider adaptations to better reach these individuals, such as improving task instructions, increasing practice trials, embedding video tutorials, or lowering the starting level of difficulty to reduce initial frustration. Despite finding the tasks widely interesting and enjoyable, fewer individuals indicated an interest in playing games at home. This difference may suggest that fewer participants would choose to engage with these tasks as a leisure activity but may be open to completing them if recommended by a member of their health care team. It is important to note that contrary to the survey wording, these tasks are not considered to be games and should not be interpreted as such. In contrast to traditional neuropsychological assessments, these measures are brief, allowing clinicians to collect a wealth of cognitive information in a relatively short period. These findings highlight the acceptability and efficiency of the battery.

Participant age was associated with 6 of the 8 tasks, with younger participants performing better than older participants. This trend has been widely observed in the cognitive literature [35-37]. Therefore, normative task data were stratified by age. Hearing sensitivity and sensitivity to mistuned, oddball, or discontinuous tones have been shown to deteriorate with age [38-40]. The slowing of processing speed with age has also been well documented [41,42]. Visuospatial working memory performance also decreases with age, potentially because of differences in chunking strategies and proactive interference [43-45]. Better performance observed with path finder, a task of visuospatial learning, may be due in part to more efficient within-task adaptability by younger individuals, a result identified previously with visuospatial tasks [43]. Other similar age-related associations have been documented in learning

XSL•FO
RenderX

paradigms [46]. The findings of this study of decreased cognitive flexibility with age are also largely supported by other studies, such as those examining task-switching capacity [47,48]. Finally, in the literature, older age has been found to be related to poorer emotion recognition, processing facial emotion, and inhibitory control [49-51]. These results likely explain why performance on tap the emotion and face to face, tasks of emotion processing and inhibitory control, was negatively associated with age; however, it is unclear why emotional face did not show an association. Given these findings, which are supported by the greater cognitive literature, normative task data were stratified by age. Therefore, normative task data were stratified by age.

Better performance on a visual learning task was observed in men, corroborating previous gender findings with learning tasks that involve spatial navigation and manipulation [52,53]. These results appear to follow previous findings, suggesting male advantage in detecting interaural time and intensity differences, complex masking tasks, and lateralization of auditory discrimination processing, although these differences are small [54,55]. With samples larger than those in this study, age- and gender-specific norms could be constructed to better characterize the individual performance.

Although previous research has demonstrated associations between performance on cognition and education tests, our task battery did not show relationships with GPA or years of education after controlling for age [56,57]. These findings may indicate that the ONAs successfully captured neural system functioning as opposed to more notion-based intelligence. In addition, only bubble pop performance was related to the occupation level after controlling for age, suggesting a potential benefit of work status.

The majority of predominantly visual ONAs were only minimally associated with one another, as expected, suggesting that they are likely to tap into unique domains of performance. Beep seeker and bubble pop, which measure sensory memory and working memory across auditory and visuospatial domains, respectively, were related. Previous research has supported this association and provided evidence for similar neural mechanisms underlying both memory systems [58,59]. In addition, 2 of the 3 tasks of emotion processing were only minimally correlated with one another. This finding is expected given the varying demands on processes of inhibition between the tasks and differences in the facial targets (eg, responding to faces with any emotion vs identifying faces with matching emotions). It is surprising that tap the emotion and emotional face were not associated, given that both tasks involved processing of facial emotions while inhibiting a prepotent response, although it is unclear whether this relationship would exist with a larger sample who performed both tasks.

Despite the recent development of various digital cognitive assessment tools, few studies have released normative data sets, thus limiting the information that can be gained from these tasks. Collecting normative data from the general community is an important and necessary first step before interpreting cognitive performance in clinical populations. The current global health crisis has arrested the ability to administer traditional in-person cognitive assessments in clinical and research settings. The development of valid and reliable assessment platforms, which can be delivered remotely, has become crucial.

Limitations and Future Work

Despite the strengths of our results, a few limitations exist that should be considered. Although 74.1% (579/781) of the sample came from the general community, a subset of participants was recruited through their attendance at a large public college institution. Therefore, young adults' performance in this sample may differ somewhat from that of the general population. Due to the limited overlap of measures that each participant completed, we were unable to evaluate whether subsets of the tasks were explained by common underlying factors to investigate construct validity. Similarly, the correlations between tasks should be interpreted with caution, given the variability in the number of individuals who were administered each task. Data on participant race and ethnicity were also not collected as part of this study. In addition, 2 individuals identified as intersex prevented our ability to reliably predict how a greater population of intersex individuals may perform on the presented tasks. Future work should aim to evaluate race and gender minorities more specifically to better understand the performance and feasibility of these populations.

Experiential cognitive assessment in patients' homes may help gain a better understanding of their true cognitive states, but it also raises the possibility that individuals will lack the attention or motivation to properly engage in testing [9]. Failure to prevent some patients from using substances or eliciting outside help during testing may also be a necessary aspect of remote testing [24,60]. In these situations, performance on a given task may not truly represent the intended cognitive domain, as it is likely to be in a distraction-limited environment. More information is also needed to understand how cognition in the general community fluctuates from shorter to more extended periods when assessed remotely.

The accessibility of these and other digital assessments provides an opportunity to integrate objective behavioral assessments directly into medical records to guide care. Given that the field of health informatics is still budding, future work should evaluate the extent to which these tasks capture unique variance and predict outcomes amidst other data.

Conclusions

This study presented the performance of individuals from the general community using a novel cognitive assessment tool that can be employed remotely. Participants found the tasks to be interactive and easy to use. The development and validation of web-based tasks such as these widely expand the accessibility of cognitive assessment to new populations such as rural groups and others with limited physical mobility and may also increase the ability of health practitioners to conduct repeated testing. Quick, easy-to-use digital assessment platforms with remote capabilities such as this may help bring the field closer to achieving more impactful patient-centered care.

XSL•FO
RenderX

Authors' Contributions

RC assisted with project development, conducted the analysis, and wrote the first draft of the manuscript. MF and BB assisted with project development, supervised the analysis, and revised and optimized further versions of the manuscript. SV led project development and revised versions of the manuscript. NG, AC, KF, and NA participated in data collection and reviewed the manuscript. All the authors contributed and have approved the final manuscript.

Conflicts of Interest

BB is a senior scientist at Posit Science, a company that produces cognitive training and assessment software. The assessments described in this study were provided for research purposes free of charge by Posit Science. SV has been a site PI on an NIH SBIR grant to Positscience Inc.

Multimedia Appendix 1
Supplementary material including online neurocognitive assessment task links and task distributions.
[DOCX File , 275 KB - jmir_v23i5e25082_app1.docx]

Multimedia Appendix 2
Participant usability feedback for online neurocognitive assessments.
[PNG File , 16 KB - jmir_v23i5e25082_app2.png]

References

1. Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. Arch Gen Psychiatry 2007 May;64(5):532-542. [doi: 10.1001/archpsyc.64.5.532] [Medline: 17485605]
2. Josiassen RC, Curry LM, Mancall EL. Development of neuropsychological deficits in Huntington's disease. Arch Neurol 1983 Dec 01;40(13):791-796. [doi: 10.1001/archneur.1983.04050120041005] [Medline: 6227312]
3. Austin MP, Mitchell P, Goodwin GM. Cognitive deficits in depression: possible implications for functional neuropathology. Br J Psychiatry 2001 Mar;178:200-206. [doi: 10.1192/bjp.178.3.200] [Medline: 11230029]
4. Raven JC, Court JH, Raven JC. Raven's progressive matrices and vocabulary scales. Oxford Psychologists Press. Oxford, England: Oxford Psychologists Press; 1998. URL: https://www.worldcat.org/title/ravens-progressive-matrices-and-vocabulary-scales/oclc/832801371 [accessed 2021-03-26]
5. Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. Am J Psychiatry 2008 Feb;165(2):203-213. [doi: 10.1176/appi.ajp.2007.07010042] [Medline: 18172019]
6. Tabert MH, Manly JJ, Liu X, Pelton GH, Rosenblum S, Jacobs M, et al. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. Arch Gen Psychiatry 2006 Aug;63(8):916-924. [doi: 10.1001/archpsyc.63.8.916] [Medline: 16894068]
7. Lee RS, Hermens DF, Naismith SL, Lagopoulos J, Jones A, Scott J, et al. Neuropsychological and functional outcomes in recent-onset major depression, bipolar disorder and schizophrenia-spectrum disorders: a longitudinal cohort study. Transl Psychiatry 2015 Apr 28;5:e555 [FREE Full text] [doi: 10.1038/tp.2015.50] [Medline: 25918992]
8. Millis SR, Rosenthal M, Novack TA, Sherer M, Nick TG, Kreutzer JS, et al. Long-term neuropsychological outcome after traumatic brain injury. J Head Trauma Rehabil 2001 Aug;16(4):343-355. [doi: 10.1097/00001199-200108000-00005] [Medline: 11461657]
9. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Clin Neuropsychol 2012;26(2):177-196 [FREE Full text] [doi: 10.1080/13854046.2012.663001] [Medline: 22394228]
10. Hafiz P, Miskowiak KW, Kessing LV, Elleby Jespersen A, Obenhausen K, Gulyas L, et al. The internet-based cognitive assessment tool: system design and feasibility study. JMIR Form Res 2019 Jul 26;3(3):- [FREE Full text] [doi: 10.2196/13898] [Medline: 31350840]
11. Resnick HE, Lathan CE. From battlefield to home: a mobile platform for assessing brain health. Mhealth 2016;2:30 [FREE Full text] [doi: 10.21037/mhealth.2016.07.02] [Medline: 28293603]
12. Monsch RJ, Burckhardt AC, Berres M, Thomann AE, Ehrensperger MM, Steiner LA, et al. Development of a novel self-administered cognitive assessment tool and normative data for older adults. J Neurosurg Anesthesiol 2019;31(2):218-226. [doi: 10.1097/ana.0000000000000510]
13. Ending disparities in mental health (EDIfy-MH). URL: https://www.nimh.nih.gov/about/organization/od/odwd/ending-disparities-in-mental-health-edify-mh.shtml [accessed 2020-12-20]

XSL•FO
RenderX

14. Hultsch DF, Strauss E, Hunter MA, MacDonald SW. Intraindividual variability, cognition, and aging. In: The Handbook of Aging and Cognition. London: Psychology Press; 2007:66.
15. Lam M, Lee J, Rapisarda A, See YM, Yang Z, Lee S, et al. Longitudinal cognitive changes in young individuals at ultrahigh risk for psychosis. JAMA Psychiatry 2018 Sep 01;75(9):929-939 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1668] [Medline: 30046827]
16. Allard M, Husky M, Catheline GL, Pelletier A, Dilharreguy B, Amieva H, et al. Mobile technologies in the early detection of cognitive decline. PLoS One 2014;9(12):- [FREE Full text] [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
17. Sbordone RJ. Limitations of neuropsychological testing to predict the cognitive and behavioral functioning of persons with brain injury in real-world settings. NeuroRehabilitation 2001;16(4):199-201. [Medline: 11790904]
18. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
19. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
20. Schwartz F, Munich RL, Carr A, Bartuch E, Lesser B, Rescigno D, et al. Negative symptoms and reaction time in schizophrenia. J Psychiatr Res 1991;25(3):131-140. [doi: 10.1016/0022-3956(91)90006-v]
21. Lövdén M, Li SC, Shing YL, Lindenberger U. Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: longitudinal data from the Berlin Aging Study. Neuropsychologia 2007 Sep 20;45(12):2827-2838. [doi: 10.1016/j.neuropsychologia.2007.05.005] [Medline: 17575988]
22. Geary EK, Kraus MF, Pliskin NH, Little DM. Verbal learning differences in chronic mild traumatic brain injury. J Int Neuropsychol Soc 2010 Mar 01;16(3):506-516. [doi: 10.1017/s135561771000010x]
23. Vinogradov S. The golden age of computational psychiatry is within sight. Nat Hum Behav 2017 Feb 8;1(2):-. [doi: 10.1038/s41562-017-0047]
24. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec;26(4):A [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
25. Biagianti B, Fisher M, Brandrett B, Schlosser D, Loewy R, Nahum M, et al. Development and testing of a web-based battery to remotely assess cognitive health in individuals with schizophrenia. Schizophr Res 2019 Jun;208:250-257 [FREE Full text] [doi: 10.1016/j.schres.2019.01.047] [Medline: 30733167]
26. Van Vleet T, Stark-Inbar A, Merzenich MM, Jordan JT, Wallace DL, Lee MB, et al. Biases in processing of mood-congruent facial expressions in depression. Psychiatry Res 2019 May;275:143-148 [FREE Full text] [doi: 10.1016/j.psychres.2019.02.076] [Medline: 30908978]
27. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019;3(1):- [FREE Full text] [doi: 10.1093/geroni/igy038] [Medline: 30619948]
28. CANTAB Mobile. URL: https://www.cambridgecognition.com/products/digital-healthcare-technology/cantab-mobile/ [accessed 2020-12-20]
29. Carter CS, Barch DM. Cognitive neuroscience-based approaches to measuring and improving treatment effects on cognition in schizophrenia: the CNTRICS initiative. Schizophr Bull 2007 Sep 17;33(5):1131-1137 [FREE Full text] [doi: 10.1093/schbul/sbm081] [Medline: 17630405]
30. Biagianti B, Fisher M, Neilands TB, Loewy R, Vinogradov S. Engagement with the auditory processing system during targeted auditory cognitive training mediates changes in cognitive outcomes in individuals with schizophrenia. Neuropsychology 2016 Nov;30(8):998-1008 [FREE Full text] [doi: 10.1037/neu0000311] [Medline: 27617637]
31. Fisher M, Holland C, Merzenich MM, Vinogradov S. Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. Am J Psychiatry 2009 Jul;166(7):805-811 [FREE Full text] [doi: 10.1176/appi.ajp.2009.08050757] [Medline: 19448187]
32. The biggest state fairs in the United States (Top 10). Readers.com. URL: https://www.readers.com/blog/biggest-state-fairs/ [accessed 2020-05-17]
33. Minnesota state fair. 2019. URL: https://assets.mnstatefair.org/pdf/20-demo-sheet.pdf [accessed 2021-03-26]
34. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol 2009 Apr;5(1):27-48. [doi: 10.1146/annurev.clinpsy.032408.153553] [Medline: 18976138]
35. Salthouse TA, Fristoe N, Rhee SH. How localized are age-related effects on neuropsychological measures? Neuropsychology 1996;10(2):272-285. [doi: 10.1037/0894-4105.10.2.272]
36. Reitan RM, Wolfson D. Influence of age and education on neuropsychological test results. Clin Neuropsychol 1995 May;9(2):151-158. [doi: 10.1080/13854049508401597]
37. Deary IJ, Der G. Reaction time, age, and cognitive ability: longitudinal findings from age 16 to 63 years in representative population samples. Aging Neuropsychol Cogn 2005 Jun;12(2):187-215. [doi: 10.1080/13825580590969235]
38. Alain C, McDonald KL, Ostroff JM, Schneider B. Age-related changes in detecting a mistuned harmonic. J Acoust Soc Am 2001 May;109(5 Pt 1):2211-2216. [doi: 10.1121/1.1367243] [Medline: 11386572]

XSL•FO
RenderX

39. Grube M, von Cramon DY, Rübsamen R. Inharmonicity detection. Effects of age and contralateral distractor sounds. Exp Brain Res 2003 Dec 3;153(4):637-642. [doi: 10.1007/s00221-003-1640-0] [Medline: 14530894]
40. Heinrich A, Schneider B. Age-related changes in within- and between-channel gap detection using sinusoidal stimuli. J Acoust Soc Am 2006 Apr;119(4):2316-2326. [doi: 10.1121/1.2173524] [Medline: 16642845]
41. Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. Age changes in processing speed as a leading indicator of cognitive aging. Psychol Aging 2007 Sep;22(3):558-568. [doi: 10.1037/0882-7974.22.3.558] [Medline: 17874954]
42. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev 1996 Jul;103(3):403-428. [doi: 10.1037/0033-295x.103.3.403] [Medline: 8759042]
43. Rowe G, Hasher L, Turcotte J. Age differences in visuospatial working memory. Psychol Aging 2008 Mar;23(1):79-84. [doi: 10.1037/0882-7974.23.1.79] [Medline: 18361657]
44. Cornoldi C, Bassani C, Berto R, Mammarella N. Aging and the intrusion superiority effect in visuo-spatial working memory. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2007 Jan;14(1):1-21. [doi: 10.1080/138255890969311] [Medline: 17164187]
45. Bo J, Borza V, Seidler RD. Age-related declines in visuospatial working memory correlate with deficits in explicit motor sequence learning. J Neurophysiol 2009 Nov;102(5):2744-2754 [FREE Full text] [doi: 10.1152/jn.00393.2009] [Medline: 19726728]
46. Howard DV, Howard Jr JH. Age differences in learning serial patterns: direct versus indirect measures. Psychol Aging 1989 Sep;4(3):357-364. [doi: 10.1037//0882-7974.4.3.357] [Medline: 2803630]
47. Taconnat L, Raz N, Toczé C, Bouazzaoui B, Sauzéon H, Fay S, et al. Ageing and organisation strategies in free recall: the role of cognitive flexibility. Eur J Cognit Psychol 2009 Mar;21(2-3):347-365. [doi: 10.1080/09541440802296413]
48. Wecker NS, Kramer JH, Hallam BJ, Delis DC. Mental flexibility: age effects on switching. Neuropsychology 2005 May;19(3):345-352. [doi: 10.1037/0894-4105.19.3.345] [Medline: 15910120]
49. Keightley ML, Winocur G, Burianova H, Hongwanishkul D, Grady CL. Age effects on social cognition: faces tell a different story. Psychol Aging 2006 Sep;21(3):558-572. [doi: 10.1037/0882-7974.21.3.558] [Medline: 16953717]
50. Wieser MJ, Mühlberger A, Kenntner-Mabiala R, Pauli P. Is emotion processing affected by advancing age? An event-related brain potential study. Brain Res 2006 Jun 22;1096(1):138-147. [doi: 10.1016/j.brainres.2006.04.028] [Medline: 16750819]
51. West R, Alain C. Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. Psychophysiology 2000 Mar;37(2):179-189. [doi: 10.1111/1469-8986.3720179]
52. Collins DW, Kimura D. A large sex difference on a two-dimensional mental rotation task. Behav Neurosci 1997;111(4):845-849. [doi: 10.1037/0735-7044.111.4.845]
53. Astur RS, Ortiz ML, Sutherland RJ. A characterization of performance by men and women in a virtual Morris water task: a large and reliable sex difference. Behav Brain Res 1998;93(1-2):185-190. [doi: 10.1016/s0166-4328(98)00019-9] [Medline: 9659999]
54. Brown CP, Fitch RH, Tallal P. Sex and hemispheric differences for rapid auditory processing in normal adults. Laterality 1999 Jan;4(1):39-50. [doi: 10.1080/713754320] [Medline: 15513103]
55. McFadden D. Sex differences in the auditory system. Dev Neuropsychol 1998 Jan;14(2-3):261-298. [doi: 10.1080/87565649809540712]
56. Ceci SJ. How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. Dev Psychol 1991;27(5):703-722. [doi: 10.1037/0012-1649.27.5.703]
57. Ceci SJ, Williams WM. Schooling, intelligence, and income. Am Psychol 1997;52(10):1051-1058. [doi: 10.1037/0003-066x.52.10.1051]
58. Pasternak T, Greenlee MW. Working memory in primate sensory systems. Nat Rev Neurosci 2005;6(2):97-107. [doi: 10.1038/nrn1603] [Medline: 15654324]
59. Bonetti L, Haumann NT, Brattico E, Kliuchko M, Vuust P, Särkämö T, et al. Auditory sensory memory and working memory skills: association between frontal MMN and performance scores. Brain Res 2018 Dec 01;1700:86-98. [doi: 10.1016/j.brainres.2018.06.034] [Medline: 29981723]
60. Tiplady B, Oshinowo B, Thomson J, Drummond G. Alcohol and cognitive function: assessment in everyday life and laboratory settings using mobile phones. Alcohol Clin Exp Res 2009;33(12):2094-2102. [doi: 10.1111/j.1530-0277.2009.01049.x] [Medline: 19740132]

Abbreviations

GPA: grade point average
MSF: Minnesota State Fair
ONA: online neurocognitive assessment
RT: reaction time
UMN: University of Minnesota

XSL•FO
RenderX

Edited by R Kukafka; submitted 18.10.20; peer-reviewed by P Hafiz, A Wright; comments to author 18.11.20; revised version received 04.01.21; accepted 16.03.21; published 06.05.21.

Please cite as:
Capizzi R, Fisher M, Biagianti B, Ghiasi N, Currie A, Fitzpatrick K, Albertini N, Vinogradov S
Testing a Novel Web-Based Neurocognitive Battery in the General Community: Validation and Usability Study
J Med Internet Res 2021;23(5):e25082
URL: https://www.jmir.org/2021/5/e25082
doi:10.2196/25082
PMID:33955839

©Riley Capizzi, Melissa Fisher, Bruno Biagianti, Neelufaer Ghiasi, Ariel Currie, Karrie Fitzpatrick, Nicholas Albertini, Sophia Vinogradov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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

Effects of an Animated Blood Clot Technology (Visual Clot) on the Decision-Making of Users Inexperienced in Viscoelastic Testing: Multicenter Trial

Sadiq Said[1]; Tadzio Raoul Roche[1]; Julia Braun[2], PhD; Micheal Thomas Ganter[3], MD; Patrick Meybohm[4], MD; Johannes Herrmann[4], MD; Kai Zacharowski[5], MD; Florian Jürgen Raimann[5], MD; Florian Piekarski[5], MD; Eva Rivas[6], MD; Manuel López-Baamonde[6], MD; Donat R Spahn[1], MD; Christoph Beat Nöthiger[1], MD; David Werner Tscholl[1], MD

[1]Institute of Anesthesiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

[2]Department of Epidemiology, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[3]Institute of Anesthesiology and Pain Therapy, Cantonal Hospital Winterthur, Winterthur, Switzerland

[4]Department of Anesthesiology, Intensive Care, Emergency, and Pain Medicine, University Hospital Wuerzburg, University of Wuerzburg, Wuerzburg, Germany

[5]Department of Anesthesiology, Intensive Care Medicine, and Pain Therapy, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt, Germany

[6]Department of Anesthesiology, Intensive Care Medicine, and Pain Therapy, Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain

Corresponding Author:
Tadzio Raoul Roche
Institute of Anesthesiology
University Hospital Zurich
University of Zurich
Raemistrasse 100
Zurich, 8091
Switzerland
Phone: 41 432530255
Email: Tadzioraoul.roche@usz.ch

Abstract

Background: Viscoelastic test–guided coagulation management has become increasingly important in assessing hemostasis. We developed Visual Clot, an animated, 3D blood clot that illustrates raw rotational thromboelastometry (ROTEM) parameters in a user-centered and situation awareness–oriented method.

Objective: This study aimed to evaluate the applicability of Visual Clot by examining its effects on users that are novices in viscoelastic-guided resuscitation.

Methods: We conducted an investigator-initiated, international, multicenter study between September 16, 2020, and October 6, 2020, in 5 tertiary care hospitals in central Europe. We randomly recruited medical students and inexperienced resident physicians without significant prior exposure to viscoelastic testing. The 7 participants per center managed 9 different ROTEM outputs twice, once as standard ROTEM tracings and once as the corresponding Visual Clot. We randomly presented the 18 viscoelastic cases and asked the participants for their therapeutic decisions. We assessed the performance, diagnostic confidence, and perceived workload in managing the tasks using mixed statistical models and adjusted for possible confounding factors.

Results: Analyzing a total of 630 results, we found that the participants solved more cases correctly (odds ratio [OR] 33.66, 95% CI 21.13-53.64; *P*<.001), exhibited more diagnostic confidence (OR 206.2, 95% CI 93.5-454.75; *P*<.001), and perceived less workload (coefficient –41.63; 95% CI –43.91 to –39.36; *P*<.001) using Visual Clot compared to using standard ROTEM tracings.

Conclusions: This study emphasizes the practical benefit of presenting viscoelastic test results in a user-centered way. Visual Clot may allow inexperienced users to be involved in the decision-making process to treat bleeding-associated coagulopathy. The increased diagnostic confidence, diagnostic certainty, reduced workload, and positive user feedback associated with this visualization may promote the further adoption of viscoelastic methods in diverse health care settings.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e27124) doi:10.2196/27124

KEYWORDS

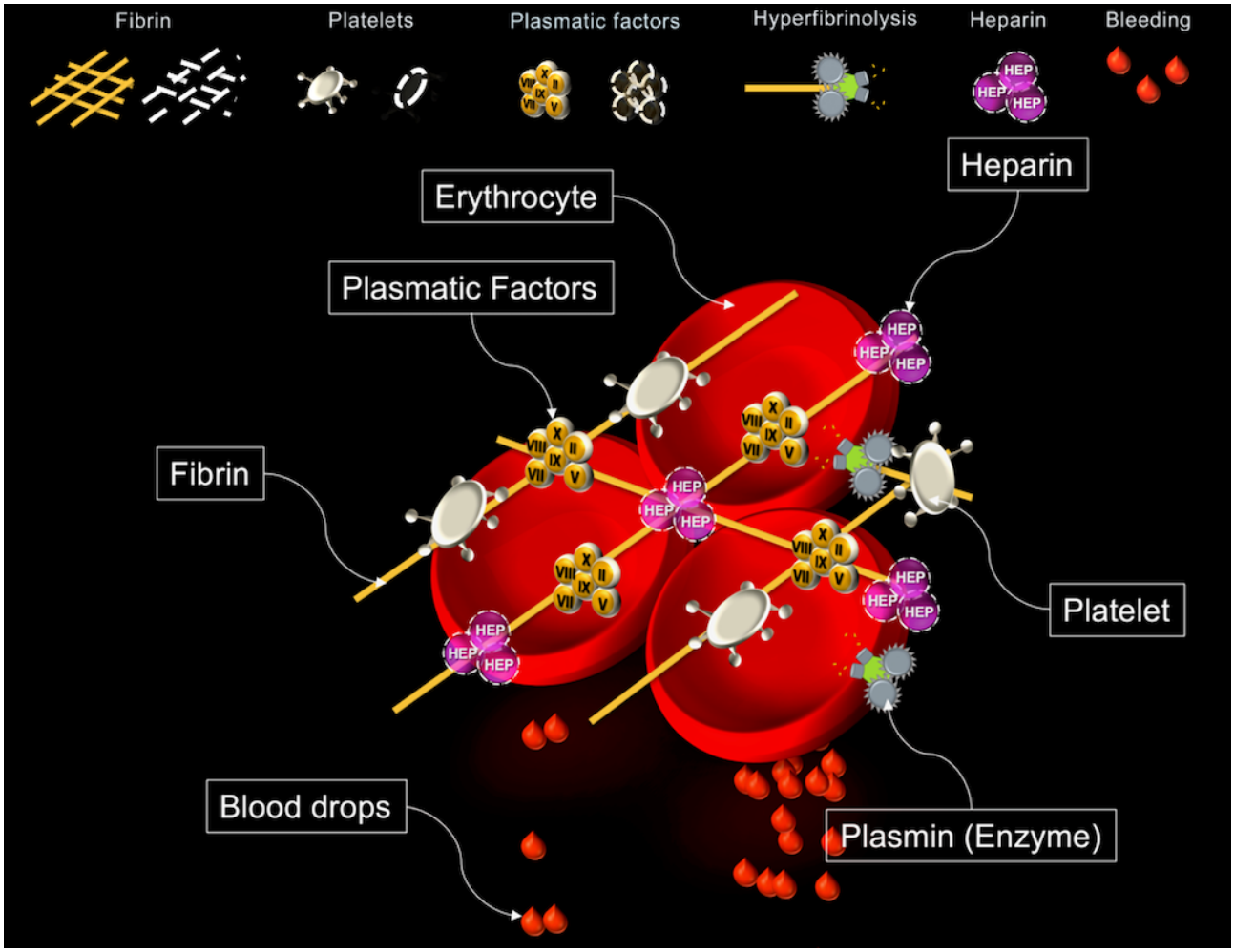
avatar technology; coagulation management; hemostasis; intuitive design; rotational thromboelastometry; user-centered design; Visual Clot; testing

Introduction

Since Hartert invented viscoelastic testing in 1948 [1] and its later clinical introduction in the 1980s [2,3], viscoelastic coagulation monitoring has become increasingly important in assessing acute bleeding in patients. To this end, several leading guidelines have proposed the use of viscoelastic-guided transfusion algorithms [4,5]. Compared to standard laboratory coagulation assays, rotational thromboelastometry (ROTEM) is faster [6,7], reduces inappropriate blood transfusions [8], and is more cost-efficient overall [9]. Further, previous studies showed that goal-directed viscoelastic hemostatic resuscitation improved patient outcomes in various different surgical specialties [9-13]. However, despite its evident importance, widespread acceptance, and increasing use, correctly interpreting ROTEM outputs remains a significant challenge for inexperienced physicians.

Hence, our research department aimed to simplify viscoelastic test outputs by developing Visual Clot technology. This animated, 3D blood clot illustrates the raw ROTEM parameters in a user-centered and situation awareness–oriented method. Visual Clot displays different coagulation components as either present or absent based on empirical ROTEM cutoff values, without making the final decision for the user. Figure 1 shows the functionality of the Visual Clot technology. In a previous, prospective, dual-center study, Visual Clot helped anesthesia and intensive care physicians in Germany and Switzerland to improve their therapeutic decisions in coagulation management [14]. In a computer-based environment, the physicians were faster, exhibited more confidence, and experienced less workload in managing the hypothetical ROTEM outputs [14]. After their initial experiences, the same physicians considered Visual Clot as intuitive, easy to learn, and useful for the decision-making process [15].

Figure 1. Example of the animated blood clot, Visual Clot. Blood drops are shown in cases of deficient hemostatic components.



In contrast to the previous research [14,15], this study aimed to evaluate the applicability of Visual Clot by examining its effects on users that are novices in viscoelastic-guided resuscitation. Without giving any instructions on analyzing or interpreting

XSL•FO
RenderX

viscoelastic results, we tested the performance of using Visual Clot compared to that of using standard ROTEM readings. We hypothesized that these inexperienced users would solve more simulated bleeding scenarios correctly, with more diagnostic confidence and less perceived workload using this avatar technology. The results of this study may support the concept of Visual Clot technology and demonstrate its potential for involving inexperienced physicians in coagulation diagnostics and management. Moreover, this study promotes the further development of user-centered, situation awareness–oriented visualization technologies.

Methods

This was an investigator-initiated, computer-based, within-subject, international multicenter study comparing standard ROTEM results with a corresponding animated viscoelastic visualization in simulated bleeding situations. We conducted this study between September 2020 and October 2020 in 5, large, tertiary care hospitals. The Cantonal Hospital Winterthur and University Hospital Zurich in Switzerland, the University Hospital Frankfurt and University Hospital Wuerzburg in Germany, and Hospital Clinic de Barcelona in Spain participated as different centers. The leading ethics committee in Zurich waived this study as it was not within the scope of the human research act (no. BASEC-Nr. Req-2020-00906). The other centers in Germany and Spain also waived ethical approval. All participants agreed in writing to the further use of obtained data for research purposes.

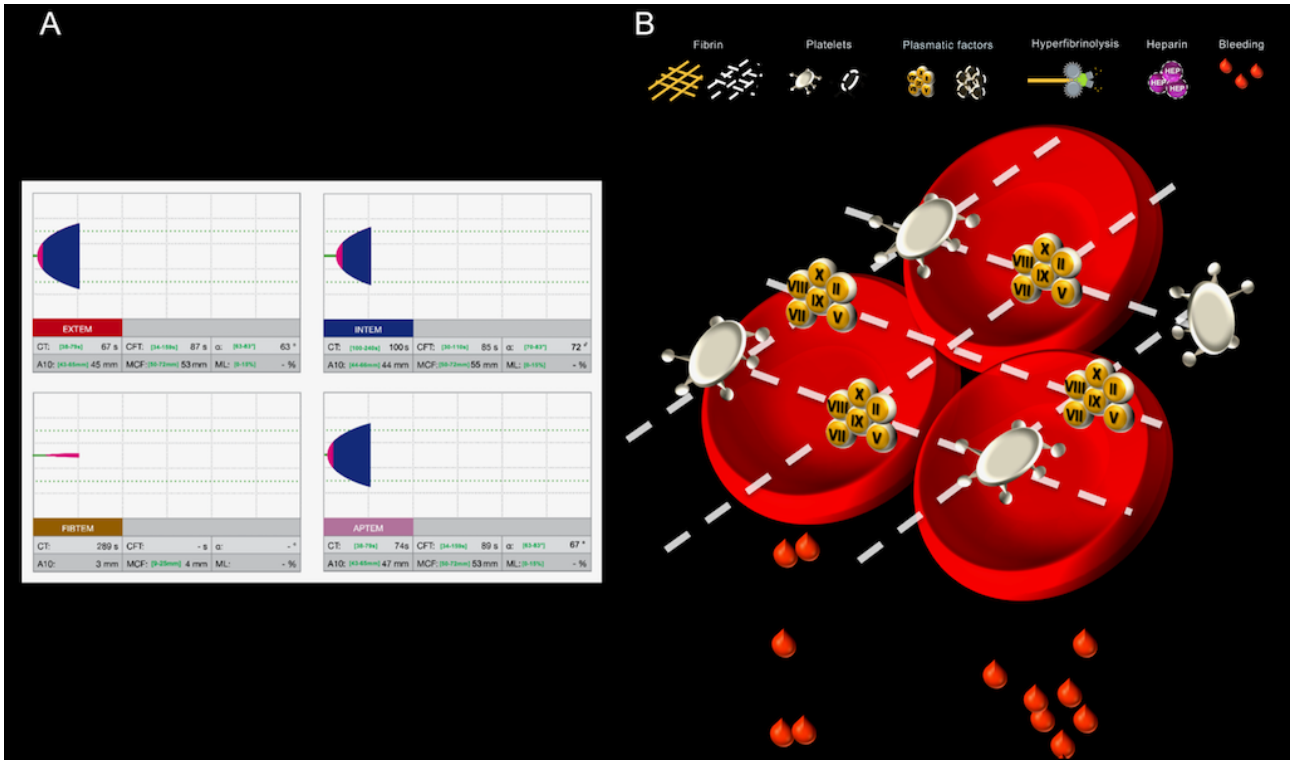
Study Procedure

For participants, we included medical students in their last or penultimate year of study and inexperienced resident physicians without significant prior exposure to viscoelastic coagulation testing. Further inclusion criteria were that they had never seen Visual Clot before and had no or minimal self-declared ROTEM skills. All participants worked in the respective hospitals. Participant selection was according to availability in clinical practice and inexperience in viscoelastic resuscitation. This subject population was completely different from that of the previous Visual Clot study [14], in which we investigated the technology's effect on experienced anesthesia and intensive care physicians.

We prepared 9 different bleeding scenarios that indicated specific coagulation disorders or a normal hemostatic state. In Multimedia Appendix 1, we provide a detailed list of all scenarios including their recommended therapeutic options. We showed each scenario twice, either as standard ROTEM readings or as the corresponding Visual Clot. Figure 2 shows an example of a standard ROTEM presentation with a corresponding Visual Clot, while Multimedia Appendix 2 shows the Visual Clot instructional video. We programmed Visual Clot according to the Zurich coagulation algorithm (Multimedia Appendix 3), which was validated in clinical practice [11]. When showing the bleeding scenarios as standard ROTEM readings, we provided the participants with the same coagulation algorithm's normal values [11]. At the beginning of the study, we first asked the participants to fill out a demographic survey on personal data, such as age, gender, and educational level. We then randomly presented the 9 different coagulation scenarios twice without any instructions on how to interpret them. We used online software [16] to randomize the sequence of these 18 scenarios, providing each participant with a unique set. We showed all ROTEM readings or the respective Visual Clot on an Apple MacBook (Apple Inc). Using exclusively the respective viscoelastic results presentation, we asked the participants to choose their targeted therapeutic recommendations from a total of 6 given answer options. It was possible that multiple therapeutic interventions were necessary for sufficient treatment. We provided the answers as checkboxes in multiple choice form using the app iSurvey (Harvest Your Data) displayed on an iPad (Apple Inc) [17]. We encouraged all participants to submit their therapeutic recommendations as quickly and accurately as possible. After each scenario, the participants rated their diagnostic confidence and perceived workload in fulfilling the given task. At the end of the study session, we asked the participants to rate 4 statements on a 5-point Likert scale (from strongly disagree to strongly agree). These statements aimed to obtain a deeper understanding of the participants' opinions about Visual Clot technology.

XSL•FO
RenderX

Figure 2. Example of a standard rotational thromboelastometry (ROTEM) presentation with a corresponding Visual Clot. Adhering to the coagulation algorithm used at the University Hospital Zurich, this hemostatic state represents fibrin deficiency.



Outcomes

The primary outcome of this study was performance, a binary outcome defined as correctly or incorrectly solved scenarios. In scenarios with multiple required therapeutic elements, we considered them to be correctly managed if the participant selected all the correct therapeutic options and no incorrect ones. As secondary outcomes of this study, we assessed the diagnostic confidence binary as unconfident or confident, and the perceived workload using the raw NASA (National Aeronautics and Space Administration) Task Load Index questionnaire. This subjective workload assessment tool has been validated in many different areas, including health care [18-23]. The raw questionnaire defines the total perceived workload as the arithmetic mean of 6 different, workload-associated subscores [23-25]. This study did not investigate the physical demand subscore, as our tasks were not physically challenging.

Statistical Analysis

As a first unadjusted analysis, we applied the McNemar test to compare the numbers of correctly and incorrectly solved cases with ROTEM and Visual Clot. We calculated mixed logistic regression models with a random intercept for each participant for the binary outcome variables, performance as correct or incorrect, and confidence as unconfident or confident. Further, we calculated a linear mixed model with a random intercept per participant for the continuous outcome regarding perceived workload as measured by the overall NASA Task Load Index scores. Apart from the variable denoting the respective viscoelastic modality (ROTEM vs Visual Clot), we adjusted all models for the center, gender, and job experience as confounders. We did not include the respective scenario as a confounding factor because the order was completely randomized for each participant and would not have consequently influenced the overall results.

As we expected similar or greater performance differences as in the previously published Visual Clot study with experienced ROTEM users, we conducted an a priori sample size calculation based on these previous results [14]. We had reported a median of 44% (317/720) correct decisions for ROTEM and 100% (720/720) correct decisions using the Visual Clot. Based on the McNemar test and assuming the proportions of correct solutions were found in the pilot study, we calculated a sample size of 15 participants to achieve a significance level of 5% and a power of 90%. As the data collection was very cost- and time efficient, we decided to include 7 participants in each center to adjust the analyses for the different centers.

We examined all data using R Version 3.6.2 (R Foundation for Statistical Computing) and created graphs using GraphPad Prism Version 9.0.0 (GraphPad Software Inc). Statistical significance was considered at a P value <.05.

Results

Between September 16, 2020, and October 6, 2020, each of the 5 study centers included 7 participants, amounting to 35 participants in total. We exposed each participant to 18 coagulation management scenarios, providing ROTEM results in 9 cases and Visual Clot in the remaining corresponding cases. No data were excluded in the final analysis. We investigated a total of 630 results, 315 per viscoelastic output modality. Regarding the participants, 49% (17 of 35) were female and none had previous contact with Visual Clot technology. Table 1 displays further study and participant characteristics. In

XSL•FO
RenderX

Multimedia Appendix 4, we provide the full statistical analysis of this study.

Table 1. Study and participant characteristics (N=35).

Characteristic	Value
Study centers, n	5
Age (years), median (IQR, range)	28 (25-32, 24-36)
Self-rated theoretical ROTEM[a] knowledge[b], median (IQR, range)	0 (0-10, 0-20)
Number of ROTEMs interpreted per year[c], median (IQR, range)	0 (0-0, 0-6])
Experience, n (%)	
Penultimate year of medical studies	3 (9%)
Last year of medical studies	10 (29%)
First year resident physician	19 (54%)
Second year resident physician	2 (6%)
Third year resident physician	1 (3%)

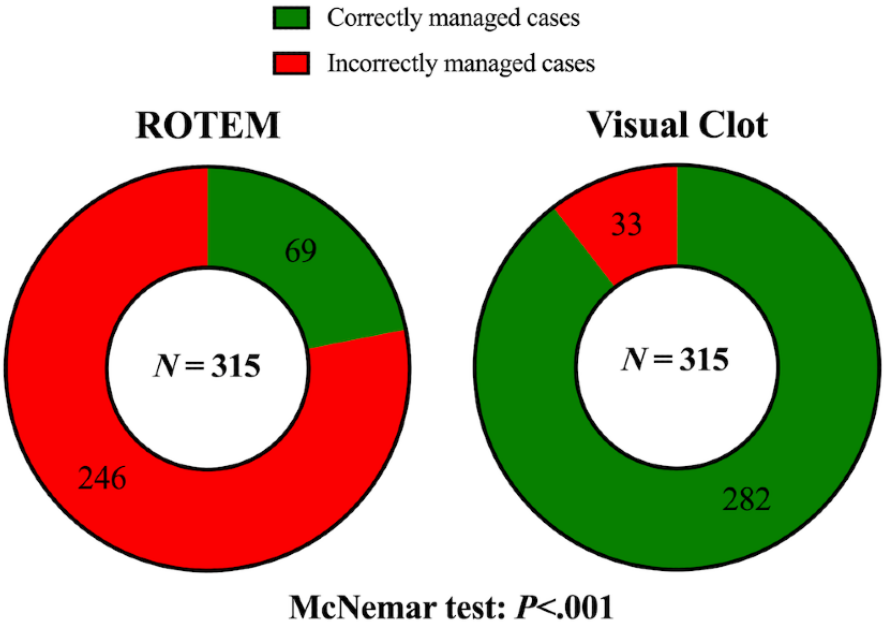
[a]ROTEM: rotational thromboelastometry.

[b]The self-rated ROTEM knowledge scale ranges from 0 (very low) to 100 (very high).

[c]The number of ROTEM interpreted per year ranges from 0 (very low) to 100 (very high).

Regarding our primary outcome, binary performance, the mixed logistic regression provided very strong evidence for a difference between the 2 viscoelastic modalities. The odds of correctly solving the case were about 33 times as high when using Visual Clot compared to when using conventional ROTEM tracings (odds ratio [OR] 33.66, 95% CI 21.13-53.64; *P*<.001). In Figure 3, we illustrate the unadjusted comparison between the viscoelastic modalities using the McNemar test. There was no significant difference in performance between the genders (OR 0.63, 95% CI 0.38-1.03; *P*=.06) or the different study centers (all *P* values >.05; Cantonal Hospital Winterthur *P*=.41; University Hospital Frankfurt *P*=.65; University Hospital Würzburg *P*=.60; Hospital Clinic de Barcelona *P*=.69). The job experience, represented by the educational level, did not differ significantly in terms of performance either (all *P*>.05; last year of studies *P*=.59; first year of residency *P*=.64; second year of residency *P*=.96; third year of residency *P*=.75).

Figure 3. Donut charts displaying the binary performance defined as the number of correctly and incorrectly solved cases using standard ROTEM results (left donut) or Visual Clot (right donut). The unadjusted analysis using the McNemar test showed very strong evidence for a difference between the 2 viscoelastic modalities (*P*<.001). N=315 per viscoelastic modality. ROTEM: rotational thromboelastometry.



For the analysis of the participants' diagnostic confidence ratings, our results showed very strong evidence in favor of Visual Clot. The odds of being diagnostically confident were about 200 times higher than when using ROTEM printouts (OR 206.2, 95% CI 93.5-454.75; *P*<.001). Regarding this outcome, no significant differences were not found for the ratings of genders (*P*=.96), centers (Cantonal Hospital Winterthur *P*=.10; University Hospital Frankfurt *P*=.42; University Hospital

XSL•FO
RenderX

Würzburg *P*=.12; Hospital Clinic de Barcelona *P*=.08), or job experiences (last year of studies *P*=.73; first year of residency *P*=.97; second year of residency *P*=.32; third year of residency *P*=.18).

Finally, mixed linear regression yielded very strong evidence for a difference between the 2 viscoelastic modalities regarding perceived workload ratings. The overall raw NASA Task Load Index scores were on average about 40 points lower if Visual Clot was used, with a coefficient of –41.63 (95% CI –43.91 to –39.36l; *P*<.001). In Figure 4, we provide the analysis of the overall perceived workload and its subscores as boxplots.

Figure 4. Boxplots representing the analysis of the participants' perceived workload after using the respective viscoelastic modality. The overall workload and its subscores were evaluated using the modified, raw NASA Task Load Index questionnaire. Low workload scores correspond to low perceived workload. The box represents the first and third quartiles, with the line indicating the median. N=315 per viscoelastic modality. The whiskers represent the 5th and 95th percentile. NASA: National Aeronautics and Space Administration; ROTEM: rotational thromboelastometry; TLX: Task Load Index.

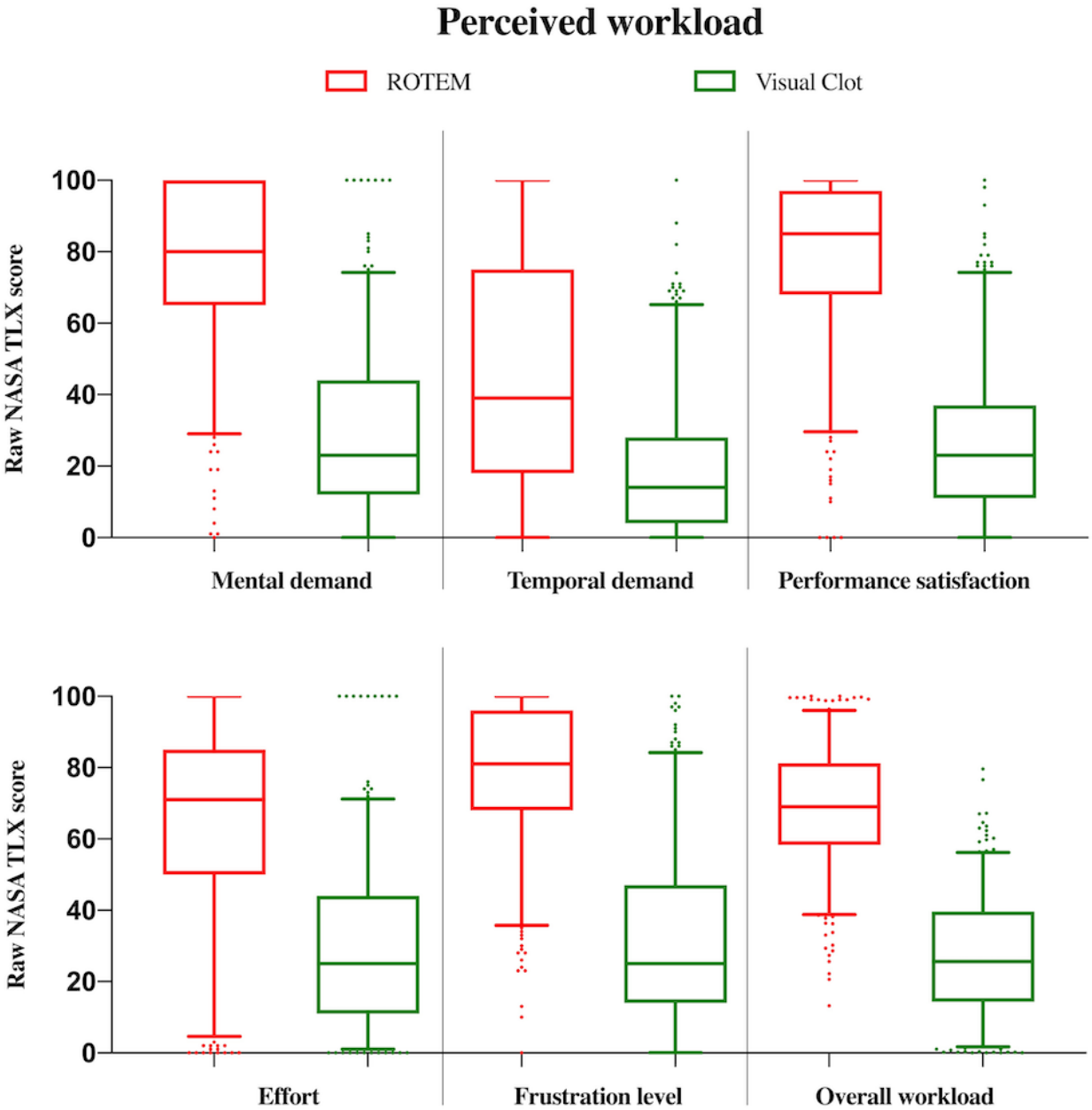


Figure 5 illustrates the 4 statements regarding the participants' opinions about Visual Clot and the results of their assessment. All following results are presented as agree (strongly agree or agree), neutral, or disagree (strongly disagree or disagree). Out of 35 participants, 34 (97%) agreed that the interpretation of Visual Clot was simple and 33 (94%) agreed that Visual Clot helped them feel better prepared to interpret viscoelastic test results. Moreover, all participants (35/34, 100%) agreed that they would use Visual Clot technology in a real bleeding situation, and 28 of 35 (80%) would want their treating physician to use Visual Clot if they were experiencing acute bleeding.

XSL•FO
RenderX

Figure 5. Graphical presentation of the participants' rated survey statements as 10 x 10 parts of whole dot plots. Results are presented as median and IQR. N=35 in each rated statement.



Discussion

This study compared 315 within-subject therapeutic decisions of 35 participants who were inexperienced in viscoelastic-guided hemostatic resuscitation and who did not receive any instructions in interpreting the result printouts. Using the animated blood clot, Visual Clot, the participants interpreted more viscoelastic test results correctly than when using the standard ROTEM tracings. Moreover, they felt more confident in their therapeutic decision and perceived less workload using Visual Clot technology. None of the participants had any previous contact with the avatar-based presentation before attending the study. We examined the results using mixed models and adjusted for different confounders.

Analysis of the participants' performance revealed that the participants selected the correct therapeutic options in over 90% (282 of 315) of the Visual Clot cases, while they made the same choices in only about 22% (69 of 315) using the corresponding ROTEM. This resulted in a relative risk ratio of about 4 and 33-times higher odds of correctly interpreting the test results when using Visual Clot technology. This positive effect persisted both across the different study centers and across the participants' educational levels. These results are even more pronounced than those in the primary Visual Clot study, where we investigated viscoelastic-experienced physicians [14]. There, the odds of a correct therapeutic decision using the avatar technology were around 22 times increased. Visual Clot appears to have significant positive effects even without any instructions for interpretation and even for users with minimal viscoelastic-guided resuscitation knowledge.

Indeed, perceived usability is a prerequisite for user acceptance of innovative technologies [26]. Both users' intuition (ie, unconscious reasoning) and the technology's characteristics, such as shape, color, and presentation of its features, directly influence how people interact with unfamiliar devices [27]. Out of 35 participants, 33 agreed that interpreting Visual Clot was simple. In a previous study investigating physicians' opinions on Visual Clot, the participants found the avatar useful, easy to learn, and intuitive [15]. Likewise, they mentioned that it allowed a faster overview of complex coagulation situations [15]. Visual Clot dichotomizes a disease state or the presence of substrates necessary for optimal clotting based on pathological, self-determinable ROTEM thresholds of an integrated, adequate coagulation algorithm. Indicating essential factors as present or absent and demonstrating them in a playful, user-centered manner reduces information complexity, leading to increased cognitive reception [28]. This binary illustration facilitates decision-making, may help to enforce local coagulation guidelines, and can reduce uncertainty in ambiguous

XSL•FO
RenderX

situations through its clear presentation. On the other hand, the standard ROTEM printouts require that the physician understands the numerical results and the data-driven tracings in the context of the coagulation algorithm's normal values to form a mental model of the current bleeding situation. This seems to cause more insecurity and incorrect therapeutic decisions that may affect the treatment of patients.

Our analysis showed a significant reduction of participants' perceived workload scores when using Visual Clot than when using traditional ROTEM results. Furthermore, our analyses showed 200 times higher odds of being confident when using Visual Clot. Again, these results are more pronounced in these novice users than in the experienced anesthesiologists and intensive care specialists of the previous Visual Clot study [14]. In large hospitals, inexperienced resident physicians may be confronted with acute bleeding situations and coagulopathy even before they possess sufficient medical training in this field. This may cause a high-pressure working environment, which is known to degrade performance [29] and lead to fatigue from perceived work overload [30]. Staff well-being directly influences the prevalence of medical errors [31], and confidence positively affects performance [32]. We should strive to minimize workload and promote the staff's diagnostic confidence to ensure better patient outcomes.

In this study, all of the participants agreed that they would use Visual Clot in a real bleeding situation. Further, 80% (28 of 35) agreed that they would want their treating physician to use this technology if they were experiencing acute bleeding themselves. It seems that the participants trust the technology and accept its application. However, we designed Visual Clot to complement the quantitative ROTEM data as a graphical representation, rather than to replace them.

This study had several limitations. Using a computer-based simulation design, we generated ROTEM printouts and corresponding Visual Clot animations that were clearly attributable to a coagulation disorder or normal hemostatic state. Viscoelastic results in real clinical bleeding may be less distinctive. Future studies are needed to confirm the results of this study in real bleeding-associated coagulopathic situations. However, simulation studies are considered an optimal environment to train and assess new methods [33]. Further, we performed this study in tertiary care hospitals in central Europe, and the results may differ elsewhere in the world. However, we consider this unlikely as all participants were novices to viscoelastic-guided management and therefore did not yet benefit from those large facilities' medical training.

This study also possesses several strengths. The analyses were adequately powered due to the a priori sample size calculation. Furthermore, the within-subject comparisons may largely rule out alternative explanations for our findings. The multicenter design and balanced participant selection across the 5 study centers minimized selection bias.

This study emphasizes the relevance of designing viscoelastic test results in a user-centered, situation awareness–oriented method. The avatar-based blood clot presentation enabled users with no or minimal knowledge in viscoelastic-guided coagulation management and without any prior training to solve almost all coagulation scenarios correctly. It further improved the participants' diagnostic confidence and reduced their perceived workload. Straightforward and confident interpretation may benefit new and experienced users in a wide range of treatment settings and promote the adoption of viscoelastic methods. The most significant benefits will likely be gained by inexperienced users and users who need to make quick decisions in stressful situations, such as on the battlefield, in spaceflight, or in the emergency room. The potential benefits of this technology and the emerging use of viscoelastic testing with its evidence-based importance justifies further investigation of Visual Clot in real clinical bleeding, with the ultimate aim of improving patient outcomes.

Acknowledgments

The authors are sincerely grateful to the staff of all participating study centers for their support in conducting this study. Further, the authors thank all study participants for their time and effort. Instrumentation Laboratory Company/Werfen Corporation had no role in the design of the study, the data collection, analysis and interpretation of data, or the decision to submit this article for publication.

This study was funded by the Institute of Anesthesiology at the University Hospital Zurich, Zurich, Switzerland.

Authors' Contributions

SS, TRR, JB, DRS, CBN, and DWT designed the study.
SS, TRR, MTG, PM, JH, KZ, ER, MLB, FJR, FP, DRS, CBN, and DWT contributed to the acquisition or interpretation of data.
SS, TRR, JB, DRS, and DWT performed data analysis.
SS, TRR, JB, MTG, PM, JH, KZ, ER, MLB, FJR, FP, DRS, CBN, and DWT drafted the manuscript or provided critical revision.
All authors gave approval of the final version of the manuscript for publication.
All authors agree to be accountable for all aspects of the work, including the accuracy and integrity of the work.

Conflicts of Interest

The academic department of DRS is receiving grant support from the Swiss National Science Foundation, Berne, Switzerland; the Swiss Society of Anesthesiology and Reanimation (SGAR), Berne, Switzerland; the Swiss Foundation for Anesthesia Research, Zurich, Switzerland; and Vifor SA, Villars-sur-Glâne, Switzerland. DRS is cochair of the ABC-Trauma Faculty, sponsored by

XSL•FO
RenderX

unrestricted educational grants from Novo Nordisk Health Care AG, Zurich, Switzerland; CSL Behring GmbH, Marburg, Germany; LFB Biomédicaments, Courtaboeuf Cedex, Franc; and Octapharma AG, Lachen, Switzerland. DRS received honoraria/travel support for consulting or lecturing from Danube University of Krems, Austria; the US Department of Defense, Washington, USA; the European Society of Anaesthesiology, Brussels, Belgium; the Korean Society for Patient Blood Management, Seoul, Korea; the Korean Society of Anesthesiologists, Seoul, Korea; the Network for the Advancement of Patient Blood Management, Haemostasis and Thrombosis, Paris, France; Baxter AG, Volketswil, Switzerland; Baxter S.p.A., Roma, Italy; Bayer AG, Zürich, Switzerland; Bayer Pharma AG, Berlin, Germany; B. Braun Melsungen AG, Melsungen, Germany; Boehringer Ingelheim GmbH, Basel, Switzerland; Bristol-Myers-Squibb, Rueil-Malmaison Cedex, France and Baar, Switzerland; CSL Behring GmbH, Hattersheim am Main, Germany and Berne, Switzerland; Celgene International II Sàrl, Couvet, Switzerland; Curacyte AG, Munich, Germany; Daiichi Sankyo AG, Thalwil, Switzerland; GlaxoSmithKline GmbH & Co. KG, Hamburg, Germany; Haemonetics, Braintree, MA, USA; Instrumentation Laboratory (Werfen), Bedford, MA, USA; LFB Biomédicaments, Courtaboeuf Cedex, France; Merck Sharp & Dohme, Kenilworth, New Jersey, USA; Octapharma AG, Lachen, Switzerland; Organon AG, Pfäffikon/SZ, Switzerland; PAION Deutschland GmbH, Aachen, Germany; Pharmacosmos A/S, Holbaek, Denmark; Photonics Healthcare BV, Utrecht, Netherlands; Pierre Fabre Pharma, Alschwil, Switzerland; Roche Diagnostics International Ltd, Reinach, Switzerland; Roche Pharma AG, Reinach, Switzerland, Sarstedt AG & Co, Sevelen, Switzerland and Nümbrecht, Germany; Schering-Plough International, Inc, Kenilworth, New Jersey, USA; Tem International GmbH, Munich, Germany; Verum Diagnostica GmbH, Munich, Germany; Vifor Pharma, Munich, Germany, Vienna, Austria, and Villars-sur-Glâne, Switzerland; Vifor (International) AG, St. Gallen, Switzerland; and Zuellig Pharma Holdings, Singapore. DRS, CBN, and DWT are the designated inventors of Visual Clot technology, for which the University of Zurich holds various patents and trademarks. The University of Zurich signed a letter of intent for a cooperation and licensing agreement with Instrumentation Laboratory Company/Werfen Corporation, Bedford, MA, USA and Barcelona, Spain. Under this and future agreements, DRS, CBN, and DWT may receive royalties. DRS, CBN, and DWT received travel support for consulting Instrumentation Laboratory, Bedford, MA, USA. CBN and DWT are designated inventors of Visual Patient technology, for which the University of Zurich holds various patents and trademarks. There are cooperation and licensing agreements with Philips Medizin Systeme Böblingen GmbH, Böblingen, Germany; Koninklijke Philips NV, Amsterdam, The Netherlands; and Philips Research/Philips Electronics Nederland BV, Eindhoven, The Netherlands. Under these agreements, CBN and DWT may receive royalties. MTG received honoraria/travel support for consulting or lecturing from Vifor Pharma of Munich, Germany; Vienna, Austria; and Villars-sur-Glâne, Switzerland. FJR received funding from HemoSonics LLC, Charlottesville, VA, USA. FJR received honoraria/travel support for consulting or lecturing from Pharma-Consult-Petersohn, Köln, Nordrhein-Westfalen, Germany; Boehringer Ingelheim, Ingelheim am Rhein, Rheinland-Pfalz, Germany; Keller Medical GmbH, Bad Soden/Ts, Hessen, Germany; and HELIOS Klinikum Krefeld, Krefeld, Germany. PM received research grants from the German Research Foundation (no. ME 3559/1-1 and ME 3559/3-1), BMBF (no. 01KG1815), the International Anesthesia Research Society, the German Society of Anaesthesiology and Intensive Care Medicine, and the European Society of Anaesthesiology; and from B. Braun Melsungen, CSL Behring, Fresenius Kabi, and Vifor Pharma for the implementation of Frankfurt‘s Patient Blood Management Program; and from Pfizer and Dr. F. Köhler Chemie GmbH for an investigator-initiated trial. PM further received honoraria for scientific lectures from Abbott GmbH & Co KG, Aesculap Academy, B. Braun Melsungen, Biotest AG, Vifor Pharma, Ferring, CSL Behring, German Red Cross/ Institute of Transfusion Medicine, HCCM Consulting GmbH, Heinen & Löwenstein, and Pharmacosmos, Siemens Healthcare; along with funding from Aktionsbündnis Patientensicherheit, the European Society of Anaesthesiology, Lohfert-Stiftung AG, Masimo Patient Safety Foundation, and MSD-Gesundheitspreis. KZ received research grants from the European Union Horizon 2020 - Research and Innovation Framework Programme (ENVISION); Aesculap Akademie GmbH, Tuttlingen, Germany; Affinites Sante, Paris, France; Ashai Kasai Pharma, Waltham, MA, USA; B. Braun AG, Melsungen, Germany; B. Braun Avitum AG, Melsungen, Germany; Bayer AG, Leverkusen, Germany; Biotest AG; Dreieich, Germany; Christian Doppler Stiftung, Wien, Österreich; CSL Behring GmbH, Marburg, Germany; Cyto Sorbents GmbH, Berlin, Germany; Edward Lifescience Corporation, Unterschleißheim, Germany; Executive Insight AG, Baar, Switzerland; Fresenius Kabi GmbH, Bad Homburg, Germany; Fresenius Medical Care, Bad Homburg, Germany; Haemonetics Corporation, Braintree, MA, USA; Hartmannbund Landesverband, Berlin, Germany; Health Advances GmbH, Zug, Switzerland; Löwenstein Medical GmbH, Bad Ems, Germany; Hexal AG, Holzkirchen, Germany; INC Research, München, Germany; Johnson & Johnson, Norderstedt, Germany; Josef Gassner, Weißenkirchen im Attergau, Austria; Maquet GmbH, Rastatt, Germany; Markus Lücke Kongress Organisation, Walsrode, Germany; Masimo International, Neuchatel, Switzerland; med Update GmbH, Wiesbaden, Germany; Medizin & Markt Gesundheitsnetzwerk, München, Germany; MSD Sharp & Dohme GmbH, Haar, Germany; Nordic Group, Hoofddorp, the Netherlands; Nordic Pharma, Ismaning, Germany; Novo Nordisk Pharma GmbH, Mainz, Germany; Pfizer Pharma GmbH, Berlin, Germany; Pharmacosmos, Wiesbaden, Germany; Ratiopharm GmbH, Ulm, Germany; Salvia Medical GmbH, Kronberg, Germany; Schering Stiftung, Berlin, Germany; Schöchl Medical Österreich, Mattsee, Austria; Serumwerke, Bernburg, Germany; Verlag für Printmedien und PR, Forum Sanitas, Leopoldshöhe, Germany; Vifor Pharma GmbH, München, Germany; Wellington, München, Germany; and Werfen, München, Germany. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
All scenarios including their recommended therapeutic options.
[PDF File (Adobe PDF File), 2082 KB - jmir_v23i5e27124_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Instructional video showing the functionality of Visual Clot.
[MP4 File (MP4 Video), 7648 KB - jmir_v23i5e27124_app2.mp4]

Multimedia Appendix 3
Visual Clot algorithm according to the local rotational thromboelastometry thresholds at the University Hospital Zurich.
[PDF File (Adobe PDF File), 152 KB - jmir_v23i5e27124_app3.pdf]

Multimedia Appendix 4
Full statistical analysis of this study.
[PDF File (Adobe PDF File), 131 KB - jmir_v23i5e27124_app4.pdf]

References

1. Hartert H. Blutgerinnungsstudien mit der thrombelastographie, einem neuen untersuchungsverfahren. Klin Wochenschr 1948 Oct;26(37-38):577-583. [doi: 10.1007/bf01697545]
2. Kang YG, Martin DJ, Marquez J, Lewis JH, Bontempo FA, Shaw BW, et al. Intraoperative changes in blood coagulation and thrombelastographic monitoring in liver transplantation. Anesth Analg 1985 Sep;64(9):888-896 [FREE Full text] [Medline: 3896028]
3. Shen L, Tabaie S, Ivascu N. Viscoelastic testing inside and beyond the operating room. J Thorac Dis 2017 Apr;9(Suppl 4):S299-S308 [FREE Full text] [doi: 10.21037/jtd.2017.03.85] [Medline: 28540073]
4. Kozek-Langenecker SA, Ahmed AB, Afshari A, Albaladejo P, Aldecoa C, Barauskas G, et al. Management of severe perioperative bleeding: guidelines from the European Society of Anaesthesiology: First update 2016. Eur J Anaesthesiol 2017 Jun;34(6):332-395. [doi: 10.1097/EJA.0000000000000630] [Medline: 28459785]
5. Spahn DR, Bouillon B, Cerny V, Duranteau J, Filipescu D, Hunt BJ, et al. The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition. Crit Care 2019 Mar 27;23(1):98 [FREE Full text] [doi: 10.1186/s13054-019-2347-3] [Medline: 30917843]
6. Haas T, Spielmann N, Mauch J, Madjdpour C, Speer O, Schmugge M, et al. Comparison of thromboelastometry (ROTEM®) with standard plasmatic coagulation testing in paediatric surgery. Br J Anaesth 2012 Jan;108(1):36-41 [FREE Full text] [doi: 10.1093/bja/aer342] [Medline: 22086509]
7. Haas T, Fries D, Tanaka KA, Asmis L, Curry NS, Schöchl H. Usefulness of standard plasma coagulation tests in the management of perioperative coagulopathic bleeding: is there any evidence? Br J Anaesth 2015 Feb;114(2):217-224 [FREE Full text] [doi: 10.1093/bja/aeu303] [Medline: 25204698]
8. Ak K, Isbir CS, Tetik S, Atalan N, Tekeli A, Aljodi M, et al. Thromboelastography-based transfusion algorithm reduces blood product use after elective CABG: a prospective randomized study. J Card Surg 2009;24(4):404-410. [doi: 10.1111/j.1540-8191.2009.00840.x] [Medline: 19583608]
9. Weber CF, Görlinger K, Meininger D, Herrmann E, Bingold T, Moritz A, et al. Point-of-Care testing. Anesthesiology 2012;117(3):531-547. [doi: 10.1097/aln.0b013e318264c644]
10. Gonzalez E, Moore EE, Moore HB, Chapman MP, Chin TL, Ghasabyan A, et al. Goal-directed hemostatic resuscitation of trauma-induced coagulopathy: a pragmatic randomized clinical trial comparing a viscoelastic assay to conventional coagulation assays. Ann Surg 2016 Jun;263(6):1051-1059 [FREE Full text] [doi: 10.1097/SLA.0000000000001608] [Medline: 26720428]
11. Stein P, Kaserer A, Sprengel K, Wanner GA, Seifert B, Theusinger OM, et al. Change of transfusion and treatment paradigm in major trauma patients. Anaesthesia 2017 Nov;72(11):1317-1326 [FREE Full text] [doi: 10.1111/anae.13920] [Medline: 28542848]
12. Mallaiah S, Barclay P, Harrod I, Chevannes C, Bhalla A. Introduction of an algorithm for ROTEM-guided fibrinogen concentrate administration in major obstetric haemorrhage. Anaesthesia 2014 Oct 07;70(2):166-175. [doi: 10.1111/anae.12859]
13. Deppe A, Weber C, Zimmermann J, Kuhn EW, Slottosch I, Liakopoulos OJ, et al. Point-of-care thromboelastography/thromboelastometry-based coagulation management in cardiac surgery: a meta-analysis of 8332 patients. J Surg Res 2016 Jun 15;203(2):424-433. [doi: 10.1016/j.jss.2016.03.008] [Medline: 27363652]
14. Rössler J, Meybohm P, Spahn DR, Zacharowski K, Braun J, Nöthiger CB, et al. Improving decision making through presentation of viscoelastic tests as a 3D animated blood clot: the Visual Clot. Anaesthesia 2020 Feb 06:14985. [doi: 10.1111/anae.14985] [Medline: 32030729]
15. Roche TR, Said S, Rössler J, Gozdzik M, Meybohm P, Zacharowski K, et al. Physicians' perceptions of a situation awareness-oriented visualization technology for viscoelastic blood coagulation management (visual clot): mixed methods study. JMIR Serious Games 2020 Dec 04;8(4):e19036 [FREE Full text] [doi: 10.2196/19036] [Medline: 33172834]
16. Urbaniak GC, Plous S. Research randomizer. Social Psychology Network. URL: http://www.randomizer.org [accessed 2020-08-01]

XSL•FO
RenderX

17. Tscholl DW, Weiss M, Spahn DR, Noethiger CB. How to conduct multimethod field studies in the operating room: the iPad combined with a survey app as a valid and reliable data collection tool. JMIR Res Protoc 2016 Jan 05;5(1):e4 [FREE Full text] [doi: 10.2196/resprot.4713] [Medline: 26732090]
18. Grier RA. How high is high? A meta-analysis of NASA-TLX global workload scores. 2015 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2015 Oct 26-30; Los Angeles. [doi: 10.1177/1541931215591373]
19. Lowndes BR, Forsyth KL, Blocker RC, Dean PG, Truty MJ, Heller SF, et al. NASA-TLX assessment of surgeon workload variation across specialties. Ann Surg 2020 Apr;271(4):686-692. [doi: 10.1097/SLA.0000000000003058] [Medline: 30247331]
20. Qiao F, Zhang R, Yu L. using NASA-Task Load Index to assess drivers' workload on freeway guide sign structures. In: American Society of Civil Engineers. 2011 Presented at: 11th International Conference of Chinese Transportation Professional (ICCTP); Aug 14 2011; Nanjing , China p. 53. [doi: 10.1061/41186(421)428]
21. Ruiz-Rabelo JF, Navarro-Rodriguez E, Di-Stasi LL, Diaz-Jimenez N, Cabrera-Bermon J, Diaz-Iglesias C, et al. Validation of the NASA-TLX score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery. Obes Surg 2015 Dec;25(12):2451-2456. [doi: 10.1007/s11695-015-1922-1] [Medline: 26459432]
22. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simul Healthc 2010 Oct;5(5):267-271. [doi: 10.1097/SIH.0b013e3181e3f329] [Medline: 21330808]
23. Said S, Gozdzik M, Roche TR, Braun J, Rössler J, Kaserer A, et al. Validation of the raw national aeronautics and space administration Task Load Index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. J Med Internet Res 2020 Sep 07;22(9):e19472 [FREE Full text] [doi: 10.2196/19472] [Medline: 32780712]
24. Hendy KC, Hamilton KM, Landry LN. Measuring subjective workload: When is one scale better than many? Hum Factors 2016 Nov 23;35(4):579-601. [doi: 10.1177/001872089303500401]
25. Nygren TE. Psychometric properties of subjective workload measurement techniques: implications for their use in the assessment of perceived mental workload. Hum Factors 2016 Nov 23;33(1):17-33. [doi: 10.1177/001872089103300102]
26. Holden H, Rada R. Understanding the influence of perceived usability and technology self-efficacy on teachers' technology acceptance. Journal of Research on Technology in Education 2011 Jun;43(4):343-367. [doi: 10.1080/15391523.2011.10782576]
27. Blackler A, Popovic V, Mahar D. Investigating users' intuitive interaction with complex artefacts. Appl Ergon 2010 Jan;41(1):72-92. [doi: 10.1016/j.apergo.2009.04.010] [Medline: 19586618]
28. Agarwal R, Karahanna E. Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Quarterly 2000 Dec;24(4):665. [doi: 10.2307/3250951]
29. Grier RA, Warm JS, Dember WN, Matthews G, Galinsky TL, Parasuraman R. The vigilance decrement reflects limitations in effortful attention, not mindlessness. Hum Factors 2003;45(3):349-359. [doi: 10.1518/hfes.45.3.349.27253] [Medline: 14702988]
30. Rose DM, Seidler A, Nübling M, Latza U, Brähler E, Klein EM, et al. Associations of fatigue to work-related stress, mental and physical health in an employed community sample. BMC Psychiatry 2017 May 5;17(1):50. [doi: 10.1186/s12888-017-1237-y]
31. Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004 Oct 28;351(18):1838-1848. [doi: 10.1056/NEJMoa041406] [Medline: 15509817]
32. Endsley M, Jones W. Wright-Patterson Air Force Base, OH: United States Airforce. 1997. URL: https://www.researchgate.net/publication/235167825_Situation_Awareness_Information_Dominance_Information_Warfare [accessed 2020-12-03]
33. Moorthy K, Vincent C, Darzi A. Simulation based training. BMJ 2005 Mar 03;330(7490):493-494. [doi: 10.1136/bmj.330.7490.493]

Abbreviations

NASA: National Aeronautics and Space Administration
OR: odds ratio
ROTEM: rotational thromboelastometry

XSL•FO
RenderX

Edited by G Eysenbach; submitted 12.01.21; peer-reviewed by R Marshall; comments to author 02.02.21; revised version received 03.02.21; accepted 11.04.21; published 03.05.21.

Please cite as:
Said S, Roche TR, Braun J, Ganter MT, Meybohm P, Herrmann J, Zacharowski K, Raimann FJ, Piekarski F, Rivas E, López-Baamonde M, Spahn DR, Nöthiger CB, Tscholl DW
Effects of an Animated Blood Clot Technology (Visual Clot) on the Decision-Making of Users Inexperienced in Viscoelastic Testing: Multicenter Trial
J Med Internet Res 2021;23(5):e27124
URL: https://www.jmir.org/2021/5/e27124
doi:10.2196/27124
PMID:33843602

©Sadiq Said, Tadzio Raoul Roche, Julia Braun, Micheal Thomas Ganter, Patrick Meybohm, Johannes Herrmann, Kai Zacharowski, Florian Jürgen Raimann, Florian Piekarski, Eva Rivas, Manuel López-Baamonde, Donat R Spahn, Christoph Beat Nöthiger, David Werner Tscholl. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2021. 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

WeChat as a Platform for Problem-Based Learning Among Hematological Postgraduates: Feasibility and Acceptability Study

Ping Luo[1], MD; Wenwen Pang[1], MD; Yingying Wang[1], PhD; Minghui Liu[1], MD; Shu Zhou[1], PhD; Shuai Liu[2], MD; Xian Zhang[1], PhD; Li Liu[1], MD; Yanan Liu[1], MD; Fuling Zhou[1], PhD

[1]Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, China

[2]Department of Hematology, Dawu County People's Hospital, Xiaogan, China

Corresponding Author:
Fuling Zhou, PhD
Department of Hematology
Zhongnan Hospital of Wuhan University
169 Donghu Road
Wuhan,
China
Phone: 86 02767814127
Email: zhoufuling@whu.edu.cn

Abstract

Background: Hematological medicine is a practical discipline that is difficult to study. Problem-based learning (PBL) is an innovative student-centered teaching method wherein students define their own learning objectives from clinically based problems. Considering that WeChat is the most popular communication app in China, we selected it as a new platform for online PBL to reduce the limitations of traditional PBL in hematology teaching.

Objective: This study aims to explore a new pedagogical method called WeChat-PBL, which is based on real micro clinical cases for postgraduates majoring in hematology and to demonstrate its feasibility and acceptability.

Methods: A total of 48 hematological postgraduates and 7 tutors participated in this study. We divided the participants into 7 groups where students can learn theoretical knowledge. After each course, the members of each group were required to complete in-class quizzes. Moreover, the students and tutors were required to fill out periodic (after each class) and overall (after each semester) evaluations.

Results: A total of 8 micro clinical cases were presented in WeChat-PBL. The average quiz score for acute myelogenous leukemia, chronic myeloid leukemia, multiple myeloma, acute promyelocytic leukemia, and lymphoma were 89.0%, 86.0%, 83.4%, 88.8%, and 77.5%, respectively. Periodic evaluations showed that both students and tutors were satisfied with the process of WeChat-PBL. The overall evaluation results showed that WeChat-PBL was able to positively impact the learning experiences of hematological postgraduates.

Conclusions: Our results indicate the feasibility and acceptability of the WeChat-PBL teaching method for postgraduates majoring in hematology.

(J Med Internet Res 2021;23(5):e16463) doi:10.2196/16463

KEYWORDS

problem-based learning; PBL; WeChat; hematology; postgraduate; education

Introduction

Postgraduate education is an important way to expand the knowledge and skills of students [1]. Today, with the continuous advancement of teaching reforms in China, the number of medical students enrolled in postgraduate degrees is increasing, especially in hematology. Hematology is a branch of medicine that explores the diagnosis, treatment, and prevention of diseases related to blood. The clinical manifestations of patients with hematological diseases are often serious and progress rapidly [2]. Both the diagnosis and treatment of hematological diseases are complex. For example, the diagnosis of leukemia, a serious disease in the hematological system, requires a number of laboratory examinations, including blood, bone marrow, immunity, molecular, and genetic analyses, which makes it difficult for students to master knowledge on blood diseases.

XSL•FO
RenderX

All of these characteristics make hematology a difficult subject to learn, and students need to change their role from passive to active. The choice of appropriate teaching methods is also important. Problem-based learning (PBL) may be one of the best methods for teaching hematology postgraduates.

The traditional pedagogy, which is teacher centered, class oriented, and examination driven, places students in a passive state of "reception" [3]. PBL is a student-centered instructional method wherein students define their own learning objectives from clinically based problems [4,5]. As an established approach, PBL has been reported to be suitable for use in graduate entry medical schools [6]. Recently, PBL has become the subject of considerable interest in postgraduate education [7]. PBL can not only cultivate postgraduates' leadership, teamwork, communication, and problem-solving abilities, which are useful for lifelong learning, but also facilitate postgraduates to become responsible for their own learning. Mindaugas and colleagues [8] found that PBL may be an effective approach for postgraduates because the PBL learning process is more demanding and self-directed, which facilitates independent and creative thinking. However, we must also admit that traditional PBL has its limitations. Generally, PBL students must get together in the classroom to discuss and share opinions. Considering the geographic and time dispersion of students, traditional PBL among hematological postgraduates is difficult to conduct. Therefore, the traditional PBL should be modified to address the physical and temporal restrictions in this cohort.

With the development of the internet in China, traditional PBL can be feasibly transferred to an online PBL setting in a virtual environment. The new mode of PBL would be constructed using modern digital technology. Thus, the time and physical restrictions of traditional PBL education would be eliminated. A previous study validated digital PBL, stating that it is as effective as traditional PBL and more effective than traditional learning in improving knowledge [9]. WeChat is one of the fastest growing mobile apps developed by Tencent, with over 697 million currently active user accounts [10], and it is the most popular platform among university students in China [11]. It has connected more than half a billion Chinese people at present. Similar to WhatsApp, WeChat permits users to send messages to an individual or a specific group in various formats, including texts, videos, voice recordings, and images [12]. WhatsApp has been introduced into medical education in colleges and universities in Western countries with great success [13,14]. In China, WeChat has also become increasingly popular as an interactive communication tool in medical education [15]. Thus , WeChat may be an appropriate tool for online PBL, which can eliminate the physical limitations of traditional PBL.

This study aims to explore a new pedagogical method called WeChat-PBL, which is based on real micro clinical cases for postgraduates majoring in hematology, and to demonstrate its feasibility and acceptability.

Methods

Participants

A total of 48 hematological postgraduates studying in Zhongnan Hospital, Wuhan University, were enrolled in our study. Clinical doctors (n=7) with 1 or more years of experience in traditional PBL teaching were assigned as tutors in the WeChat-PBL groups. Our new mode of PBL was constructed on the basis of the WeChat app. WeChat is a popular app that is available on Android, iPhone, and Windows. It is supported by Wi-Fi, 4G, and 5G data networks. Using WeChat, students can communicate with each other anywhere and at any time [16]. All students who participated in this project had their own mobile phones with the WeChat app installed. All were familiar with the practical aspects of WeChat in the PBL context. We randomly divided the participants into 7 groups. Each group comprised 6 or 7 students and 1 tutor. WeChat-PBL was used from August to October 2020. Informed consent was provided to the participants and signed prior to beginning the study. The study was approved by the Medical Council of Wuhan University.

The WeChat-PBL Pedagogy Method

Micro clinical cases including clinical characteristics and relevant questions were uploaded to the WeChat-PBL group 3 days before class. The WeChat group included a tutor and 6 or 7 postgraduate students. Then, the 7 members were divided into 2 groups to create their own Microsoft PowerPoint (Microsoft Corp) according to the micro medical cases the teacher presented by reading books, reviewing the literature, or asking others. In the PBL class, the representative of each group uploaded their PowerPoint and showed it to the others using the video conferencing function in WeChat. Members in the WeChat-PBL group could then pose questions about the selected case and discuss various issues by sending text, images, voice recordings, videos, or documents. As a person who facilitates and motivates learning in the group, the tutor summarized and explained the questions. Owing to the functions of WeChat, all the messages could be easily and instantly read, making WeChat-PBL highly efficient.

The WeChat-PBL Student Examination System

PBL is a learner-centered instructional method by which students learn content and thinking strategies [17]. Self-regulated learning skills play an important role in learning efficiency [18]. Some students are highly self-disciplined, whereas others are not. To help students with poor self-discipline, examination is necessary. All participants were required to complete a quiz after each class, graded by the tutor, and the average score was calculated based on the quiz result. The tutor visually displayed the students' progress through the WeChat-PBL group and monitored students who performed poorly. This assessment was conducted immediately after each class.

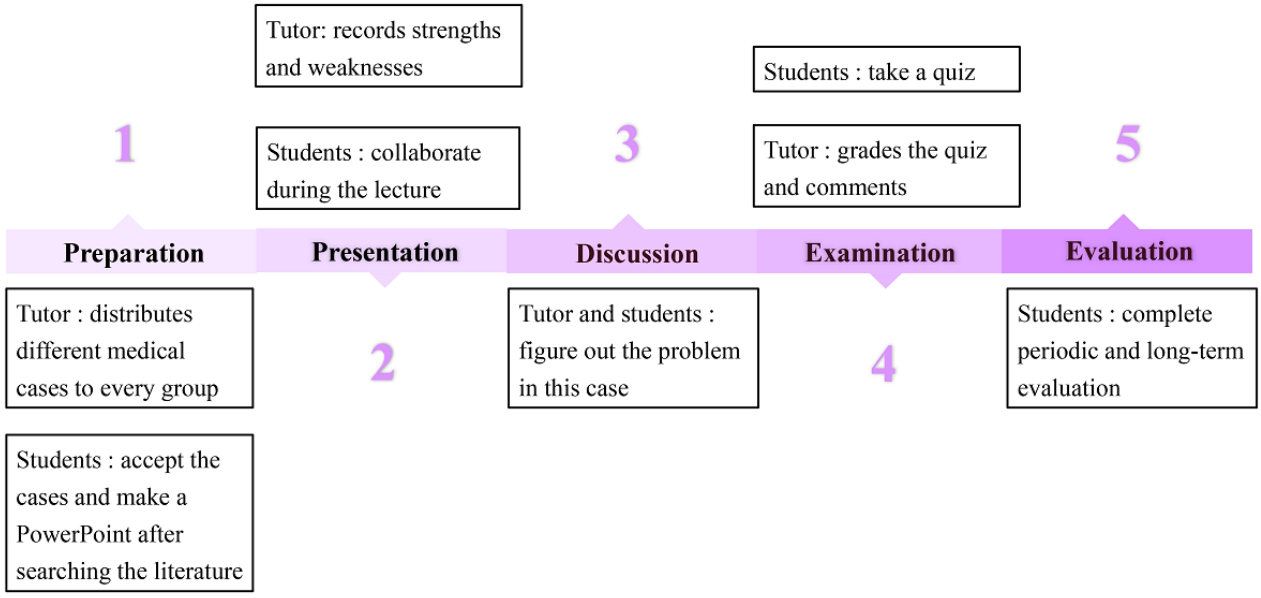
Evaluating WeChat-PBL

Curriculum evaluation is important to ensure the quality of education [19], and this also applies to WeChat-PBL. The key to evaluating WeChat-PBL is to assess whether postgraduates are effectively gaining knowledge, improving skills, and

XSL•FO
RenderX

developing scientific research abilities. The evaluations were divided into 2 sections: periodic evaluation of the WeChat-PBL process and an overall evaluation of the whole WeChat-PBL project. The students and tutors were required to complete the periodic evaluation after each PBL class (eg, evaluation of students' participation in creating PowerPoints and performance in the WeChat-PBL group). The overall evaluation of WeChat-PBL was conducted in the form of a group discussion after each semester. The questions were designed in accordance with previous studies [20,21] and with the specific features of this study. The questions displayed in the *Results* section were discussed by all tutors who participated in this research to ensure its quality. The basic flowchart of the WeChat-PBL mode is shown in Figure 1.

Figure 1. Basic flowchart of the WeChat-PBL teaching mode.



Analysis

Microsoft Excel 2013 (Microsoft Corp) was used to perform the data analysis. The average score for each micro clinical case (ie, disease) was calculated by adding up all the quiz scores and divided by the total number of participants involved in the case. Percentages were used to describe the overall evaluation results and was calculated by the number of participants who agreed with each item divided by the total number of participants (N=48). The content of the periodic evaluations within the WeChat-PBL groups was analyzed qualitatively.

Results

PBL Cases

In our study, 8 real clinical cases were presented in WeChat-PBL. A total of 5 different kinds of hematological diseases—acute myelogenous leukemia, chronic myeloid leukemia, multiple myeloma, acute promyelocytic leukemia, and lymphoma—were studied. Every WeChat-PBL group comprised 3 classes. It took 6-10 days to finish 1 PBL case.

Student Examination

The average quiz score for the 5 kinds of hematological diseases were as follows: acute myelogenous leukemia, 89.0%; chronic myeloid leukemia, 86.0%; multiple myeloma, 83.4%; acute promyelocytic leukemia, 88.8%; and lymphoma, 77.5%. The majority of students (41/48, 85.4%) communicated actively within the groups, whereas 7 communicated less actively in the beginning. The tutor spoke with them privately, inquired about the reason behind the reduced participation, and encouraged them to actively take part in the discussions. Finally, all students became active in the group discussions.

Evaluation of WeChat-PBL

Periodic evaluation results showed that every students played an active part in making their PowerPoint. The tutors were satisfied with the performance of the students and reported that students exhibited obvious improvement in both theoretical study and scientific research abilities.

The overall evaluation was conducted in the form of a group discussion. The results of this discussion are presented in Table 1. Most students had a better understanding of PBL (41/48, 85.4%). Almost all students were satisfied with their performance in WeChat-PBL, including completing the tasks assigned to them on time (44/48, 91.7%), applying prior knowledge to solve problems (48/48, 100.0%), and communicating ideas with group members effectively (42/48, 87.5%) The majority (44/48, 91.7%) believed that their clinical skills and scientific research abilities were improved through the use of WeChat-PBL. In addition, most of the students reported that group members and the tutor offered timely feedback (45/48, 93.8%) and held the view that the WeChat-PBL teaching mode was reasonable (42/48, 87.5%) and better than traditional PBL (39/48, 81.2%). The majority of students thought that WeChat was an effective app for online PBL for postgraduates majoring in hematology (39/48, 81.2%). In terms of the quality of the micro clinical cases presented in WeChat-PBL, the majority of students thought that it was valuable and interesting (43/48, 89.6%). Not surprisingly, almost all students felt that the role of the tutor was very important in

XSL•FO
RenderX

the PBL group setting (46/48, 95.8%). However, only 30 out of 48 participants (62.5%) felt free to pose questions in the WeChat-PBL group.

Table 1. Results of the overall evaluation of the WeChat-PBL teaching mode.

Number	Question	Respondents, n (%)	
		Agree	Disagree
1	Did you have a better understanding of PBL[a]?	41 (85.4)	7 (14.6)
2	Did you enjoy the discussions in WeChat-PBL?	41 (85.4)	7 (14.6)
3	Did you complete the tasks assigned to you on time?	44 (91.7)	4 (8.3)
4	Did you apply prior knowledge to solve problems?	48 (100.0)	0 (0.0)
5	Did you feel free to pose questions?	30 (62.5)	18 (37.5)
6	Did you actively take part in the discussions?	34 (70.8)	14 (29.2)
7	Were you able to communicate ideas with your group members effectively?	42 (87.5)	6 (12.5)
8	Do you think that the micro clinical cases presented in WeChat-PBL are valuable and interesting?	43 (89.6)	5 (10.4)
9	Did your clinical skills and scientific research abilities improve through WeChat-PBL?	44 (91.7)	4 (8.3)
10	Did your group members and tutor offer timely feedback?	45 (93.8)	3 (6.2)
11	Do you think the teaching mode of WeChat-PBL is reasonable?	42 (87.5)	6 (12.5)
12	Do you think the teaching mode of WeChat-PBL is better than that of traditional PBL?	39 (81.2)	9 (18.8)
13	Did your tutor consistently show enthusiasm with PBL?	48 (100.0)	0 (0.0)
14	Do you think that the tutor played an important role in student learning in the PBL group?	46 (95.8)	2 (4.2)
15	Do you think that WeChat is an effective app for online PBL for postgraduates majoring in hematology?	39 (81.2)	9 (18.8)

[a]PBL: problem-based learning.

Discussion

Principal Findings

The new PBL mode based on WeChat was designed to promote the abilities of hematological postgraduates, including clinical reasoning, team skills, and metacognition. This mode of online PBL successfully eliminated the physical and temporal limitations of traditional PBL in postgraduate teaching. In our study, the average quiz score for acute myelogenous leukemia, chronic myeloid leukemia, multiple myeloma, and acute promyelocytic leukemia was very high. The majority of students reported that their clinical skills and scientific research abilities improved through WeChat-PBL and thought that WeChat was an effective app for online PBL for postgraduates majoring in hematology, indicating the acceptance and effectiveness of the WeChat-PBL teaching mode.

In this study, almost all students felt that the role of the tutor was very important in the PBL group. A previous study has also validated that the tutor plays an important role in student learning in the PBL group [22]. Dissatisfaction with PBL has arisen owing to the teaching capabilities and enthusiasm of the facilitators [23,24]. Barrows and colleagues [25] expressed the view that the tutor should facilitate and motivate learning rather than serve as a source of knowledge. However, other studies support the idea that the best tutors are those with both the ability to promote learning and present clinical content [26]. Dolmans et al [27] have reported that the tutor's acts not only have an impact on the productiveness but also the effectiveness of the PBL group's work. To help tutors become more effective in PBL tutorials, continuous training that can facilitate tutors' reflections of their own development as teachers [28,29] and assist them in shifting from a traditional lecturing role to a multifaceted role as a mentor, coach, model, and guide [30] is needed.

The majority of data on tutor training has mainly focused on general moderation techniques for tutors [31]. In 2005, Azer [32] presented a group of challenges faced by PBL tutors and 12 tips for successful group facilitation. Tutors can read books, educational reviews, and research articles and regularly record teaching experiences by using reflective journals to achieve the essence of these tips in teaching. Moreover, PBL has 7 sequential steps, including case presentation, problem definition, brainstorming, generating hypotheses, defining learning goals, self-study, and synthesis [21]. Tutors should cultivate themselves according to these steps.

Eight micro clinical cases were presented in WeChat-PBL, and 89.6% of students found these cases to be valuable and interesting. Because the cases are brief and interesting,

XSL•FO
RenderX

WeChat-PBL is time-efficient, effective, and interesting to the learner and teacher. Our study involved real-life cases and the actual experiences of health professionals, which is helpful to enhance the relevance of the subject matter. Professional knowledge is integrated with clinical presentations. Student learning, therefore, is associated with real-life situations.

PBL is a learner-centered instructional pedagogy, in which students play a central role in their learning [17]. Thus, several competencies including knowledge acquisition, practical skills, and professional attitudes are important for them to become active, cooperative, and self-directed [33]. Assessment of competence is vital for students to identify and respond to their own learning needs, providing insights into their actual performance. Examination drives learning, and it affects not only what students learn but also how they learn [34]. In our study, we assessed the students by evaluating their performance in the WeChat-PBL group, completion of the work, and a quiz after each class. We observed an obvious correlation between the examination scores of students who participated more actively than others.These data highlight the importance of effective assessment systems for monitoring student progress in the PBL curricula.

Of note, our WeChat-PBL teaching mode has several obvious advantages compared with the traditional PBL. First, WeChat is the most popular app in China [35]. It is a very convenient way for students to communicate. By uniting traditional classroom education with the WeChat app, students can learn actively at any time, both before and after class. Second, the frequent notifications in the WeChat group establish periodic contact between students, which encouraged student feedback on the course and facilitated student-student and student-teacher interactions. Third, more teamwork can be encouraged through WeChat-PBL than traditional PBL given that students can communicate easily with the help of the WeChat app. Fourth, the micro clinical cases were uploaded to the WeChat-PBL group before class. Students were able to prepare for the session in advance, thus cultivating the learner in an efficient, goal-directed manner.

Conclusion

A great number of changes have taken place in the instructional methods used to teach medical students. Hematological medicine is a practical discipline that is difficult to master. In our study, we demonstrated the feasibility and acceptability of the WeChat-PBL teaching mode for postgraduates majoring in hematology. The new PBL mode is time-saving and convenient. Additionally, it provides a suitable platform for sharing the latest information and educational resources. It emphasizes interoperable, interactive, effective, and participatory teaching styles. Despite the limited sample size in our study, our results indicate that WeChat-PBL is applicable to postgraduates majoring in hematology. In the future, further comprehensive, large-scale, and high-quality studies should be conducted to confirm our findings and thus promote the applications of this new teaching mode for hematological postgraduates.

Acknowledgments

This study was supported by the Wuhan University Postgraduate Tutor Education Innovation Project in 2019 (201942), the Wuhan University Medical Department Teaching Research Project in 2019 (2019056), the Wuhan University MOOC Course Construction Project in 2020 (413200007), the Wuhan University Postgraduate Quality Curriculum Construction Project in 2020 (413200068), and the 2020 National Steering Committee for Postgraduate Education for Medical Degrees (A-YXC20200202-06).

Conflicts of Interest

None declared.

References

1. Hanley A. Venereology at the polyclinic: postgraduate medical education among general practitioners in England, 1899-1914. Med Hist 2015 Apr;59(2):199-221 [FREE Full text] [doi: 10.1017/mdh.2015.3] [Medline: 25766540]
2. Chen X, Chen B, Li X, Song Q, Chen Y. Mutual benefit for foreign medical students and Chinese postgraduates: A mixed team-based learning method overcomes communication problems in hematology clerkship. Biochem Mol Biol Educ 2017 Mar 04;45(2):93-96 [FREE Full text] [doi: 10.1002/bmb.20997] [Medline: 27801537]
3. Tu M, Yu C, Wu L, Li T, Kwan C. Dental and medical students' perspectives on early exposure to PBL in Taiwan. J Dent Educ 2012 Jun;76(6):746-751. [Medline: 22659703]
4. Aldayel AA, Alali AO, Altuwaim AA, Alhussain HA, Aljasser KA, Bin Abdulrahman KA, et al. Problem-based learning: medical students' perception toward their educational environment at Al-Imam Mohammad Ibn Saud Islamic University. Adv Med Educ Pract 2019;10:95-104 [FREE Full text] [doi: 10.2147/AMEP.S189062] [Medline: 30881170]
5. Rui Z, Rong-Zheng Y, Hong-Yu Q, Jing Z, Xue-Hong W, Chuan Z. Preliminary investigation into application of problem-based learning in the practical teaching of diagnostics. Adv Med Educ Pract 2015;6:223-229 [FREE Full text] [doi: 10.2147/AMEP.S78893] [Medline: 25848334]
6. O Doherty D, Mc Keague H, Harney S, Browne G, McGrath D. What can we learn from problem-based learning tutors at a graduate entry medical school? A mixed method approach. BMC Med Educ 2018 May 04;18(1):96 [FREE Full text] [doi: 10.1186/s12909-018-1214-2] [Medline: 29728090]
7. Kilroy DA. Problem based learning. Emerg Med J 2004 Jul;21(4):411-413 [FREE Full text] [doi: 10.1136/emj.2003.012435] [Medline: 15208220]

XSL•FO
RenderX

8. Stankunas M, Czabanowska K, Avery M, Kalediene R, Babich SM. The implementation of problem-based learning in health service management training programs. Leadersh Health Serv (Bradf Engl) 2016 Oct 03;29(4):392-401. [doi: 10.1108/LHS-04-2015-0010] [Medline: 27707149]
9. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital Problem-Based Learning in Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945 [FREE Full text] [doi: 10.2196/12945] [Medline: 30816846]
10. Wang J, Gao F, Li J, Zhang J, Li S, Xu G, et al. The usability of WeChat as a mobile and interactive medium in student-centered medical teaching. Biochem. Mol. Biol. Educ 2017 Jun 05;45(5):421-425. [doi: 10.1002/bmb.21065] [Medline: 28581156]
11. Sun Z, Zhu L, Liang M, Xu T, Lang J. The usability of a WeChat-based electronic questionnaire for collecting participant-reported data in female pelvic floor disorders: a comparison with the traditional paper-administered format. Menopause 2016 Aug;23(8):856-862. [doi: 10.1097/GME.0000000000000690] [Medline: 27326820]
12. Pang H. WeChat use is significantly correlated with college students' quality of friendships but not with perceived well-being. Heliyon 2018 Nov;4(11):e00967 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00967] [Medline: 30519660]
13. Oyewole BK, Animasahun VJ, Chapman HJ. A survey on the effectiveness of WhatsApp for teaching doctors preparing for a licensing exam. PLoS One 2020;15(4):e0231148 [FREE Full text] [doi: 10.1371/journal.pone.0231148] [Medline: 32240259]
14. Hussain I, Khan RA, Mahboob U, Latif MZ, Waheed K. Comparison of effectiveness of WhatsApp and Facebook as learning tool for the students of a medical college. J Pak Med Assoc 2020 Oct;70(10):1705-1708 [FREE Full text] [doi: 10.5455/JPMA.35071] [Medline: 33159737]
15. Zeng F, Deng G, Wang Z, Liu L. WeChat: a new clinical teaching tool for problem-based learning. Int J Med Educ 2016 Apr 25;7:119-121 [FREE Full text] [doi: 10.5116/ijme.5708.e5c4] [Medline: 27111920]
16. Tu S, Yan X, Jie K, Ying M, Huang C. WeChat: An applicable and flexible social app software for mobile teaching. Biochem Mol Biol Educ 2018 Sep;46(5):555-560 [FREE Full text] [doi: 10.1002/bmb.21170] [Medline: 30369037]
17. Wang Q, Li H, Pang W. From PBL tutoring to PBL coaching in undergraduate medical education: an interpretative phenomenological analysis study. Med Educ Online 2016;21:31973 [FREE Full text] [doi: 10.3402/meo.v21.31973] [Medline: 27396900]
18. Demirören M, Turan S, Öztuna D. Medical students' self-efficacy in problem-based learning and its relationship with self-regulated learning. Med Educ Online 2016;21:30049 [FREE Full text] [doi: 10.3402/meo.v21.30049] [Medline: 26987386]
19. Galukande M, Katamba A, Kiguli S, Kiguli-Malwadde E, Kijjambu S, Sewankambo N. Problem based learning: tutors' views 5 years after implementation at a sub-Saharan University. Afr Health Sci 2015 Mar;15(1):261-268 [FREE Full text] [doi: 10.4314/ahs.v15i1.34] [Medline: 25834557]
20. Zhang W, Li Z, Li Z. WeChat as a Platform for Problem-Based Learning in a Dental Practical Clerkship: Feasibility Study. J Med Internet Res 2019 Mar 19;21(3):e12127. [doi: 10.2196/12127]
21. Bosse HM, Huwendiek S, Skelin S, Kirschfink M, Nikendei C. Interactive film scenes for tutor training in problem-based learning (PBL): dealing with difficult situations. BMC Med Educ 2010 Jul 06;10:52 [FREE Full text] [doi: 10.1186/1472-6920-10-52] [Medline: 20604927]
22. Lyberg-Åhlander V, Lundskog M, Hansson K. Experiencing the role of PBL tutor. Clin Linguist Phon 2014;28(1-2):36-46. [doi: 10.3109/02699206.2013.816371] [Medline: 23895224]
23. Burgess A, Bleasel J, Haq I, Roberts C, Garsia R, Robertson T, et al. Team-based learning (TBL) in the medical curriculum: better than PBL? BMC Med Educ 2017 Dec 08;17(1):243 [FREE Full text] [doi: 10.1186/s12909-017-1068-z] [Medline: 29221459]
24. Burgess A, Ayton T, Mellis C. Implementation of team-based learning in year 1 of a PBL based medical program: a pilot study. BMC Med Educ 2016 Mar 04;16:49 [FREE Full text] [doi: 10.1186/s12909-016-0550-3] [Medline: 26846425]
25. Maudsley G. Roles and responsibilities of the problem based learning tutor in the undergraduate medical curriculum. BMJ 1999 Mar 06;318(7184):657-661 [FREE Full text] [doi: 10.1136/bmj.318.7184.657] [Medline: 10066213]
26. Papinczak T, Tunny T, Young L. Conducting the symphony: a qualitative study of facilitation in problem-based learning tutorials. Med Educ 2009 Apr;43(4):377-383. [doi: 10.1111/j.1365-2923.2009.03293.x] [Medline: 19335580]
27. Dolmans DHJM, Wolfhagen IHAP. Complex interactions between tutor performance, tutorial group productivity and the effectiveness of PBL units as perceived by students. Adv Health Sci Educ Theory Pract 2005 Aug;10(3):253-261. [doi: 10.1007/s10459-005-0665-5] [Medline: 16193404]
28. Hendry GD. Problem-based learning tutors' conceptions of their development as tutors. Med Teach 2009 Mar;31(2):145-150. [doi: 10.1080/01421590802146026] [Medline: 19330673]
29. Bollela VR, Gabarra MHC, da Costa C, Lima RCP. Students and tutors' social representations of assessment in problem-based learning tutorials supporting change. BMC Med Educ 2009 Jun 07;9:30 [FREE Full text] [doi: 10.1186/1472-6920-9-30] [Medline: 19500408]
30. Hitchcock MA, Mylona ZH. Teaching faculty to conduct problem-based learning. Teach Learn Med 2000;12(1):52-57. [doi: 10.1207/S15328015TLM1201_8] [Medline: 11228868]

XSL•FO
RenderX

31. Van Berkel HJM, Dolmans DHJM. The influence of tutoring competencies on problems, group functioning and student achievement in problem-based learning. Med Educ 2006 Aug;40(8):730-736. [doi: 10.1111/j.1365-2929.2006.02530.x] [Medline: 16869917]
32. Azer SA. Challenges facing PBL tutors: 12 tips for successful group facilitation. Med Teach 2005 Dec;27(8):676-681. [doi: 10.1080/01421590500313001] [Medline: 16451886]
33. Dolmans DHJM, De Grave W, Wolfhagen IHAP, van der Vleuten CP. Problem-based learning: future challenges for educational practice and research. Med Educ 2005 Jul;39(7):732-741. [doi: 10.1111/j.1365-2929.2005.02205.x] [Medline: 15960794]
34. Al Kadri HM, Al-Moamary MS, van der Vleuten C. Students' and teachers' perceptions of clinical assessment program: A qualitative study in a PBL curriculum. BMC Res Notes 2009 Dec 23;2:263 [FREE Full text] [doi: 10.1186/1756-0500-2-263] [Medline: 20030842]
35. Dong Y, Wang P, Dai Z, Liu K, Jin Y, Li A, et al. Increased self-care activities and glycemic control rate in relation to health education via Wechat among diabetes patients: A randomized clinical trial. Medicine (Baltimore) 2018 Dec;97(50):e13632 [FREE Full text] [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]

Abbreviations

PBL: problem-based learning

Edited by R Kukafka; submitted 01.10.19; peer-reviewed by R de Leeuw, C Mather; comments to author 31.08.20; revised version received 08.11.20; accepted 16.03.21; published 25.05.21.

Please cite as:
Luo P, Pang W, Wang Y, Liu M, Zhou S, Liu S, Zhang X, Liu L, Liu Y, Zhou F
WeChat as a Platform for Problem-Based Learning Among Hematological Postgraduates: Feasibility and Acceptability Study
J Med Internet Res 2021;23(5):e16463
URL: https://www.jmir.org/2021/5/e16463
doi:10.2196/16463
PMID:34032573

©Ping Luo, Wenwen Pang, Yingying Wang, Minghui Liu, Shu Zhou, Shuai Liu, Xian Zhang, Li Liu, Yanan Liu, Fuling Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Transitioning to Web-Based Learning in Basic Life Support Training During the COVID-19 Pandemic to Battle the Fear of Out-of-Hospital Cardiac Arrest: Presentation of Novel Methods

Katarzyna Naylor[1], PhD; Kamil Torres[1], MBA, PhD, MD

Department of Didactics and Medical Simulation, Medical University of Lublin, Lublin, Poland

Corresponding Author:
Katarzyna Naylor, PhD
Department of Didactics and Medical Simulation
Medical University of Lublin
Chodźki 6
Lublin, 20-093
Poland
Phone: 48 814485922
Email: katarzynanaylor@umlub.pl

Abstract

Ongoing training in the area of basic life support aims to encourage and sustain the willingness to act in out-of-hospital cardiac arrest situations among first aiders. The contribution of witnesses and first aiders has diminished rapidly, as suspicion associated with the COVID-19 pandemic has risen. In this paper, we present teaching methods from the medical education field to create a new teaching-learning process for sustaining the prehospital involvement of first aiders and encourage new first aiders. The most important benefit—improving outcomes—can be achieved by introducing a variety of teaching-learning methods and formative assessments that provide participants with immediate feedback to help them move forward in the basic life support course. The new reality of web-based learning that has been introduced by the pandemic requires an innovative approach to traditional training that involves techniques and methods that have been proven to be useful in other fields.

(J Med Internet Res 2021;23(5):e27108) doi:10.2196/27108

KEYWORDS

COVID-19; web-based training; basic life support; formative assessment; out-of-hospital cardiac arrest; web-based learning; web-based education; first aid; medical education; life support; transition; outcome; formative

Introduction

Due to public health concerns in the wake of 2020, which were caused by the novel COVID-19, face-to-face contact in medical training was promptly substituted with remote teaching. The teaching-learning process was moved to the houses of participants through a variety of learning management systems and videoconferencing services [1]. Similarly, all hands-on resuscitation training in the form of in-person, hands-on sessions was stopped [2]. Nevertheless, the European Resuscitation Council (ERC) as well as the International Liaison Committee on Resuscitation (ILCOR), in their educational update in April 2020 on teaching during a pandemic, highlighted the importance of sustaining sudden cardiac arrest training in some form, despite the modified conditions for knowledge transfer. The ILCOR and national societies underlined the significance of continuing education to improve resuscitation knowledge, certain skills, and, most importantly, patient-centered care to sustain and encourage the willingness to act during out-of-hospital cardiac arrest (OHCA) situations [3,4]. Complementary basic life support (BLS) teaching methods involving computer-based and video-based e-learning or the gamification approach have been proven to be useful in the area of BLS [5,6]. However, the cost-effectiveness and standardization of the training delivered via these educational methods should also be considered [7,8].

The long-term retention of BLS competencies is essential and outweighs skills performance during teaching sessions, according to published evidence [9,10]. Such pieces of evidence have opened a new door in the era of BLS training. Additionally, as BLS training is conducted in each health care professional undergraduate medical curriculum, there have been additional demands for finalizing modules that were planned for educational months, which changed due to the COVID-19 pandemic. The Polish Ministry of Health, together with the Ministry of Education, suspended face-to-face education from March 12, 2020, onward [11]. Therefore, the opportunity for

XSL•FO
RenderX

launching a web-based BLS course arose (e-BLS) in response to new conditions.

To expand people's knowledge about distant and web-based possibilities for learning in the area of BLS, we present a proposal for a newly designed e-learning course about BLS competencies. This course was developed to (1) enable participants to understand and cultivate the necessary competencies in each of the OHCA domains and (2) promote the implementation of the steps required for helping people in need of BLS. The course design entails a novel teaching methodology that has been extensively researched in the medical education field.

Methods

The Goal Of Training and Its Objectives

A multimodal, participant-centered, interactive, web-based course for addressing the unique challenges of sudden cardiac arrest that a layperson may face daily was developed. The overarching goals of the course are to (1) enable participants to understand and cultivate the necessary competencies in each of the OHCA domains and (2) promote the implementation of the steps required for helping people in need of BLS. Specific objectives were determined by analyzing the ERC guidelines update for the COVID-19 pandemic. This was done to set up BLS course learning objectives and the needs assessment of the participants (Table 1).

XSL•FO
RenderX

Table 1. Objectives, instructional methods, and implementation strategies for each basic life support (BLS) competency session.

Session topic (number of sessions; time of each session)	Objectives	Instructional design and implementation
Safety (2; 45 minutes)	• Personal safety and the safety of people in need of BLS in cardiac arrest situations during the COVID-19 pandemic (hand hygiene, the donning and doffing of gloves and masks, types of masks, methods for approaching a person in need of BLS, and scene safety)	• Video-based training [12]: Students review the instructional videos on the e-learning platform, produce instructional videos to present their skills, and receive feedback from the faculty on their performance.
BLS (3; 45 minutes)	• Standard algorithms and modifications resulting from pandemic (methods that do not involve the look, listen, and feel technique and continuous, chest compression–only CPR[a] without mouth-to-mouth ventilation)	• Video-based training and decision trees [13]: After familiarizing themselves with the assigned materials and conducting video discussions, students are presented with selected sudden cardiac arrest recognition and management scenarios related to BLS. Students have 5 attempts to solve each of the two decision trees and receive immediate feedback.
AED[b] (3; 45 minutes)	• Description of the equipment, equipment use, and where and how to find equipment during an out-of-hospital cardiac arrest • Using an AED as a nonaerosol-producing step	• Video-based training and decision trees: After familiarizing themselves with the assigned materials and conducting video discussions, students are presented with selected sudden cardiac arrest recognition and management scenarios related to BLS. Students have 5 attempts to solve decision trees and receive immediate feedback. • Simple game scenario [6]: Participants implement the BLS/AED algorithm in a simple virtual environment, and the player or student has to save a person by applying CPR actions. • Find your AED: Participants find the closest device to their place of residence by using the Staying Alive app [14].
pBLS[c] and FBAO[d] (3; 45 minutes)	• Standard algorithms and modifications resulting from the pandemic	• Video-based learning and decision trees: After familiarizing themselves with the assigned materials and conducting video discussion, students are presented with selected pediatric sudden cardiac arrest recognition and management scenarios related to pBLS. Students have 5 attempts to solve 1 decision tree and receive immediate feedback. • Instructor-led, live practice session on Zoom about FBAO algorithms for infants and chest compressions for infants: Participants use available toys resembling a newborn.
Special circumstances leading to sudden cardiac arrest, such as anaphylaxis, heart attacks, strokes, diabetes, drowning, hypothermia, burns, and seizures (3; 45 minutes)	• Definitions, standard algorithms, and modifications resulting from the pandemic	• Peer assessments [15] and teacher assessments of the presentations of assigned topics are prepared in a group of 3 people and recorded by participants.
Review (4; 45 minutes)	• Review of all of the topics	• The Script Concordance Test [16,17]: This includes 8 BLS scenarios accompanied by questions concerning possible next steps (Multimedia Appendix 1). Each of the questions is supplemented by a new piece of information concerning the sudden cardiac arrest health issue being considered. Students use a 3-point Likert scale to determine whether and to what extent the new piece of information influences further evaluations.

[a]CPR: cardiopulmonary resuscitation.

[b]AED: automated external defibrillator.

[c]pBLS: pediatric basic life support.

[d]FBAO: foreign body airway obstruction.

Participants: Target Group

The course aims to provide an introduction to BLS and strengthen the elements of the chain of survival. No professional BLS experience is required. However, the course can also be treated as a refresher course for maintaining participants' motivation to assist in OHCA situations and providing an update on BLS training to health professionals. Therefore, the target

XSL•FO
RenderX

population includes students of medical faculties (medicine, dentistry, nursing, midwifery, biomedicine, paramedics, public health, and dieticians), first aiders with initial first aid knowledge, lay rescuers who are enrolled in a first aid course, and health care providers who want to refresh their knowledge.

The e-learning modules have been embedded in the curriculum of students in all medical faculties of the local medical university, as first aid is obligatory in all curricula. The decision trees and the Script Concordance Test (SCT) have been prepared in Polish and English language for Polish and foreign students. The course is planned for 20 teaching hours and takes place in an academic environment.

To target other professional groups and voluntary first aiders, a massive, open, web-based course will be established to enable their participation and access to the proposed activities.

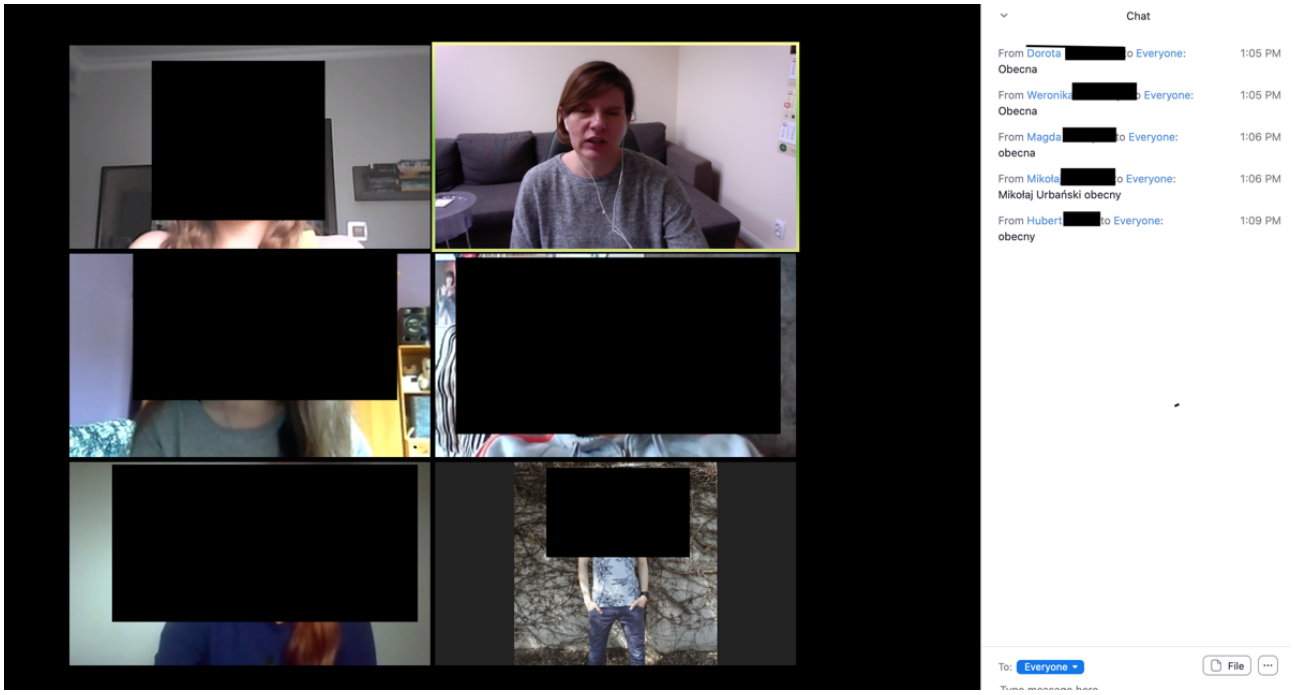
Evaluation

The course will be assessed on the basis of participants' achievements on the final, summative, multiple-choice question test; the SCT; and decision trees as well as data from an anonymous survey based on the Utrecht Seminar Evaluation questionnaire by Spruijt et el [18]. The authors made the tool available upon our request.

Learning Management Systems and Videoconferencing Services

Based on the advice of experienced researchers, two platforms that are well known to participants were chosen to host the course—Zoom (Zoom Video Communications Inc; Figure 1) and Moodle (Figures 2 and 3)—to alleviate uncertainty and distrust among the students [19]. Zoom enabled live meetings with participants for discussing a given topic, whereas Moodle, a web-based guiding platform, was used to store all of the information on the course meetings and requirements and the organization, all of the materials and resources assigned to topics, and the links to the formative tasks described in this paper. To adjust to web-based classes, participants received detailed instructions on the use of both platforms. The decision trees' software constituted another university software; however, links to each task appeared on Moodle and were assigned to a given subject (BLS, BLS/automated external defibrillator [AED], or pediatric BLS), and detailed instructions on how to access the task as well as the rules of engagement were provided.

Figure 1. An anonymized screen shot from an individual Zoom meeting.



XSL•FO
RenderX

Figure 2. A general overview of an introductory Moodle course website.

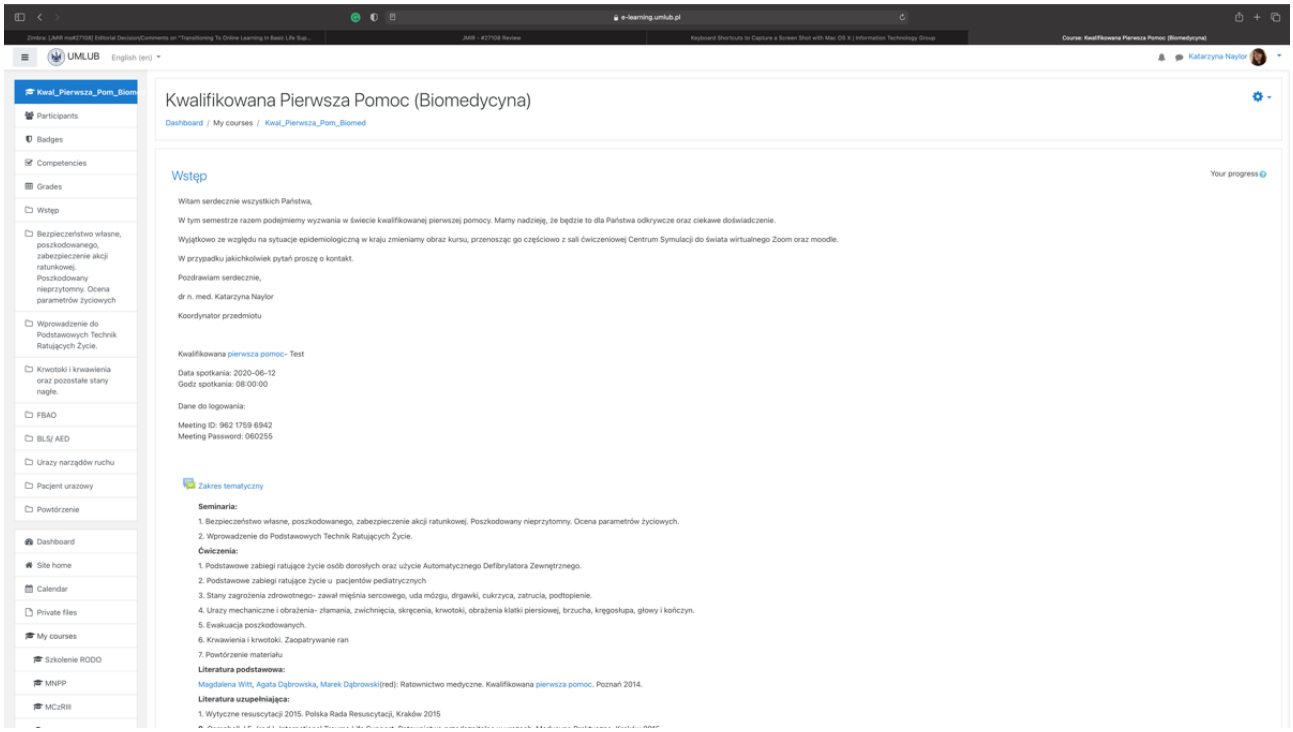
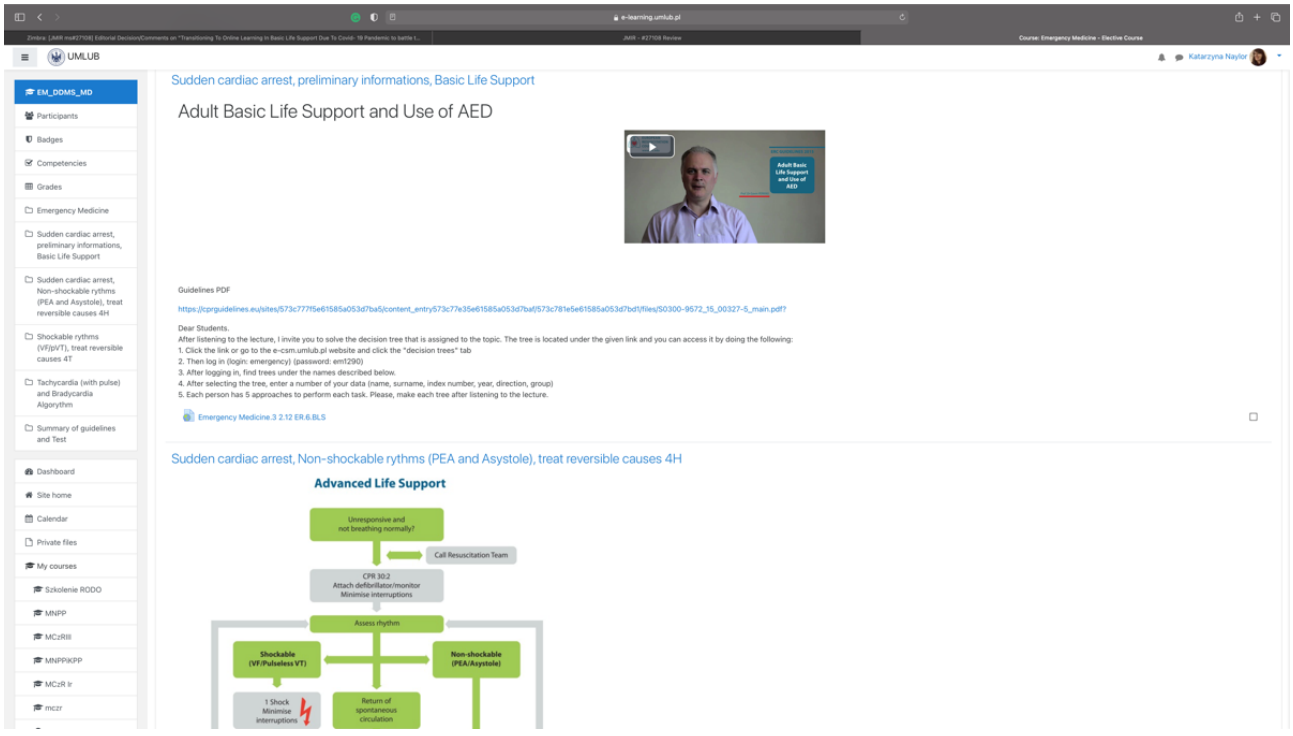


Figure 3. A screenshot of another course on Moodle and a decision tree for instruction. AED: automated external defibrillator; PEA: pulseless electrical activity; VF: ventricular fibrillation; VT: ventricular tachycardia.



The educational game for engaging with BLS/AED drills was also located on a different server. A freely available resource was used [20]. Elements of the game were scored, and participants posted their obtained results on the Moodle forum.

Another type of software that was used during the course was a newly developed testing system—the Testing Centre for Medical Exams (Centrum Medycznych Egzaminów Testowych [CMET]) [21]. The final summative assignment was held on CMET. The CMET used the following 3-level system for constructing the summative assessment (multiple-choice questions): (1) teachers input the questions into the system; (2) specialists in the area reviewed the questions; and (3) the course coordinator accepted, edited, or sent back the questions (with comments) to their primary creator. This was done to ensure the quality of the provided tasks and generate a more complex task (Figures 4 and 5).

Before the final assignment, the participants receive access to a practice test to confirm their login details and familiarize

XSL•FO
RenderX

themselves with the system's construction. During the test, each student is individually timed during their attempt and receives immediate feedback on the questions alongside their results after they finalize their attempt.

Figure 4. A screenshot of a login page to the CMET website. CMET: Centrum Medycznych Egzaminów Testowych.

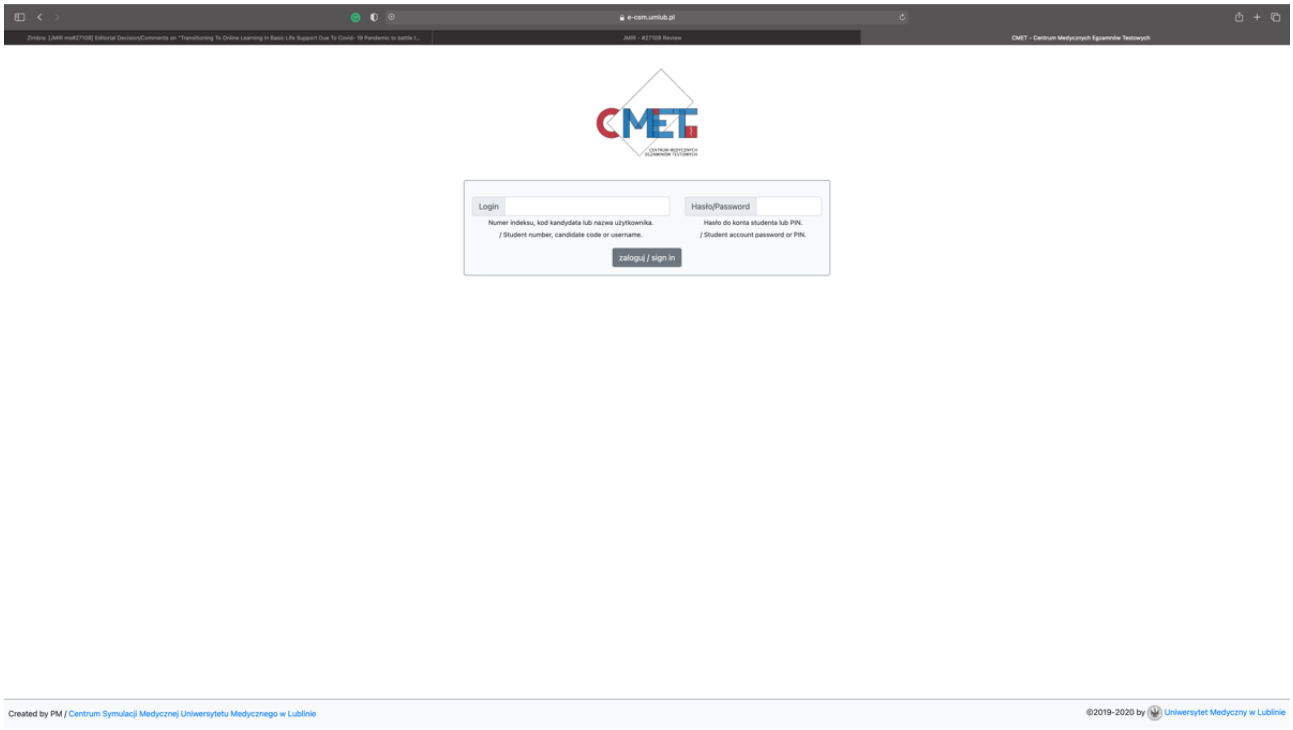
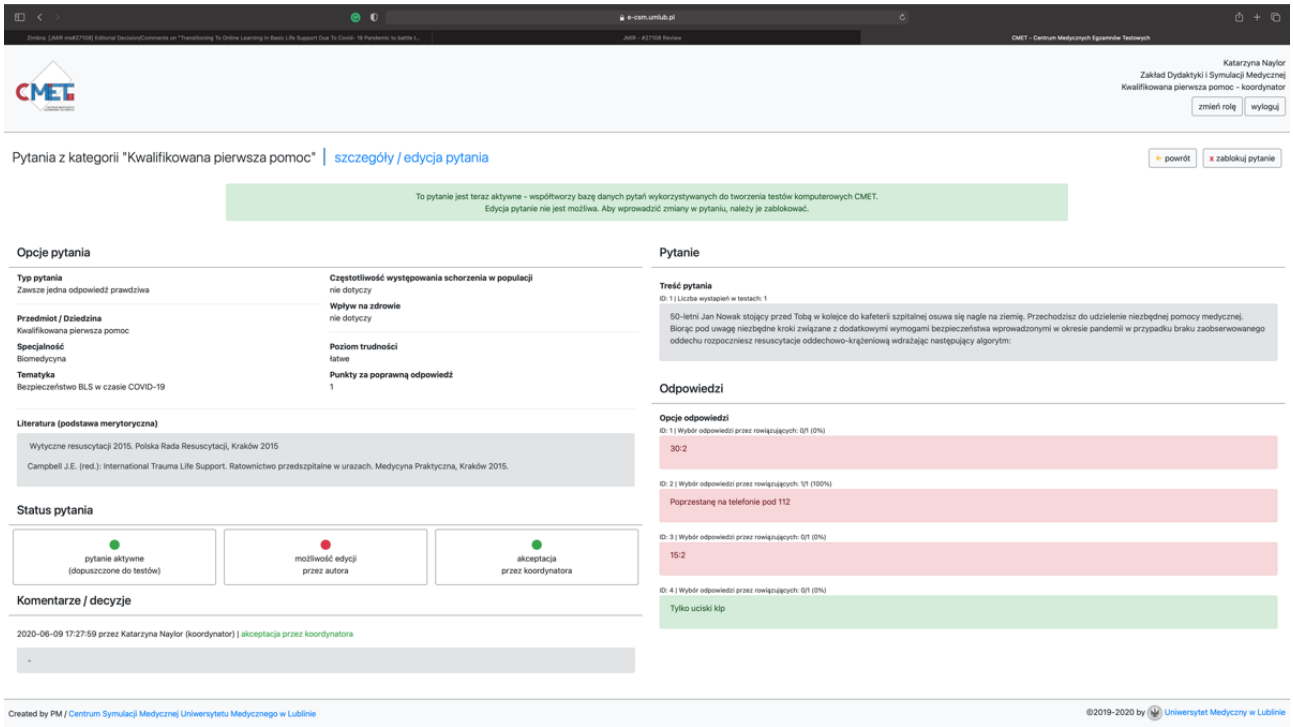


Figure 5. A screenshot of a coordinator account page for a single question on the CMET website. CMET: Centrum Medycznych Egzaminów Testowych.



Discussion

Content and Facilitation

The implemented assessment points support learning and correspond with the implemented content. Our assessments points focus on formative assessment; they are not used as a tool for passing judgment and making pass-fail decisions but as a tool for assisting with and supporting learning [22]. Formative assessments are also known as assessments for learning or learning-oriented assessments. Such assessments aim to provide feedback on understanding and help participants move forward in the course (Table 1).

XSL•FO
RenderX

Novel Assessment Formats That Promote Knowledge Acquisition

Learning is promoted through the formative tasks in the outlined BLS web-based course. Participants need to be consciously analyzing and applying gathered "intel" on a given subject in order to benefit from the experience. All of the proposed types of formative assessments in our course aim to produce feedback on performance to advance the learning of participants [23].

The American Heart Association (AHA) has implemented a video-based approach (a teaching approach centered on using videos) that allows AHA course participants to practice BLS techniques while watching videos (ie, practice-while-watching method) during AHA courses. The AHA also provides the option of an e-learning module, in which participants familiarize themselves with the content of videos and textual content before participating in hands-on sessions [24]. Additionally, the ERC has focused on implementing the Peyton approach (ie, a 4-step method) during their hands-on sessions [25]. Therefore, the focus of BLS courses was placed on face-to-face training, and web-based materials were only used to prepare for face-to-face meetings.

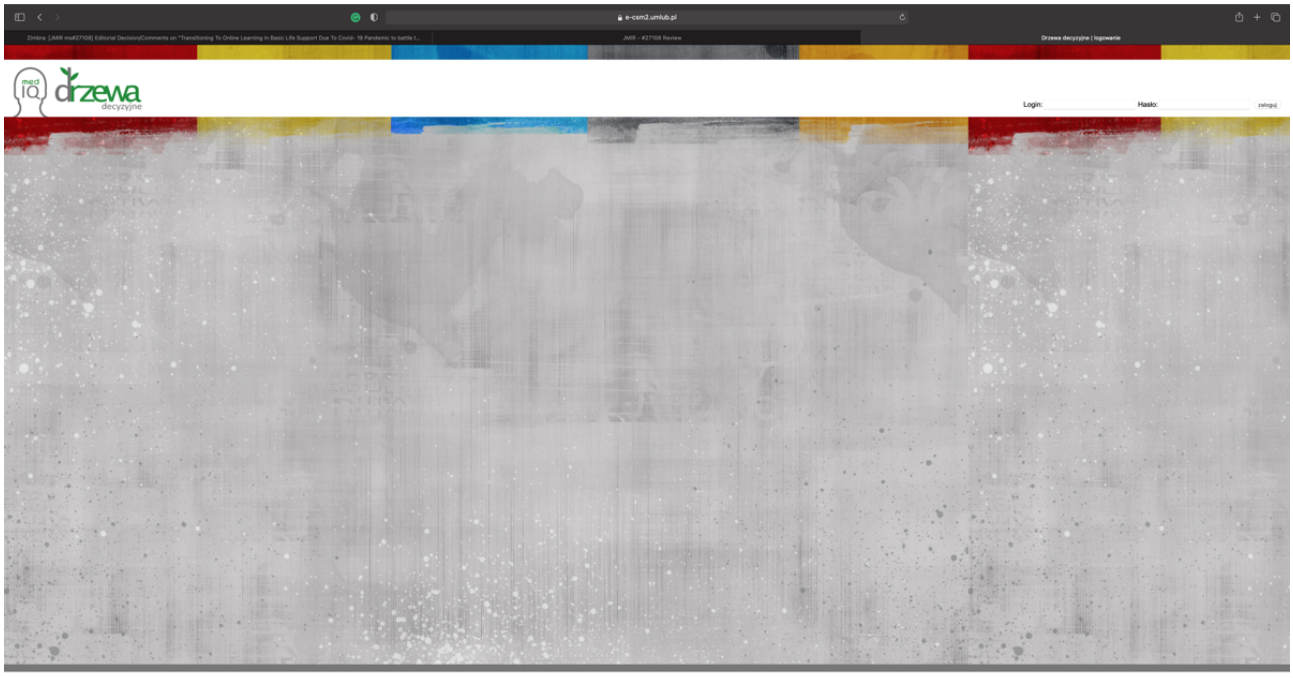
The SCT

The SCT as an assessment method for evaluating clinical reasoning that has several advantages in BLS training. Students tackle a genuine case to demonstrate their ability to incorporate new data into the information on the provided scenario [16,17]. They can compare their reasoning and conduct related discussions, thereby allowing them to learn from each other. This is especially important in emergency situations and OHCA situations, in which quick and timely decisions are required. Additionally, the SCT goes beyond pure fact checking; it requires logical thinking and knowledge application based on procedural knowledge (Multimedia Appendix 1).

Decision Trees

Decision trees are based on the assumption that machine learning algorithms are helpful in the medicine field, as errors can have a dire consequence. Decision trees have been used in the decision-making process in clinical settings [13,26]. Nevertheless, the existing system allows first aiders to tackle the complexities and uncertainties of OHCA. Existing decision tress allow for the implementation of knowledge at the more critical level of the Millers pyramid in BLS competency training (Multimedia Appendix 2; Figure 6).

Figure 6. A screenshot of a login website for accessing the decision trees.



Gamification

The BLS/AED game was designed to complement BLS teaching and refresh skills in a unique, enjoyable way [6]. The game presents an OHCA scenario by using simple graphics and takes players to a virtual environment where they make choices based on the steps of the BLS/AED algorithm (Multimedia Appendix 2). The game is educational and is a unique medium for knowledge transmission; such games have been proven to promote active learning, help with solving clinical problems, and help learners gain experience in a risk-free environment [6]. At the end of the game, the participants of our course will receive feedback on their performance, and they will have an opportunity to undergo another attempt (Multimedia Appendix 3).

Conclusions

There is a large variety of web-based first responder and BLS courses. Although the new reality resulting from a pandemic and teaching and learning during a pandemic can have difficulties, they also create new possibilities. Enabling continuous BLS training with appropriate tools, including a

XSL•FO
RenderX

variety of didactic methods and assessment formats, can facilitate an uninterrupted process of learning for battling the fear of out-of-hospital resuscitation during the COVID-19 pandemic. However, the modifications in distant BLS learning are at an early stage, and there is a need for robust research that determines (1) their association with participant outcomes; (2) their impact on OHCA; and (3) whether such methodologically diverse learning is cost-effective. Addressing these issues will provide further insight into the role and effectiveness of new technologies and their potential impact on acquiring and sustaining BLS competencies.

Possible Implications for Practice

First, during the pandemic, the OHCA and BLS curricula require more care for nurturing and supporting competencies in the area of public health. Second, an integrated web-based program requires the combination of modern technology and formative assessments that are dedicated to developing critical thinking and decision-making skills that encourage people to take action in OHCA situations. Third, appropriate modern resources are an integral part of creating modern curricula for BLS that encourage people to take action in OHCA situations.

Limitations

One of the limitations of our research may be that a lot of time is required to complete all of the formative activities. An increase in the amount of time devoted to a given training course puts an additional burden on course participants. However, well-thought-out activities that mirror real-life situations should encourage participants to spend more time in tackling such activities.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The proposed Script Concordance Test.
[DOCX File , 86 KB - jmir_v23i5e27108_app1.docx]

Multimedia Appendix 2
A recording of an exemplary decision tree used during the course.
[MOV File , 112401 KB - jmir_v23i5e27108_app2.mov]

Multimedia Appendix 3
A recording of the game used during the course. Audio commentary in English is provided.
[MP4 File (MP4 Video), 3995 KB - jmir_v23i5e27108_app3.mp4]

References

1. Schlegel E. Designing online courses: 12 tips for health professions educators. MedEdPublish 2020;9(1):117 [FREE Full text] [doi: 10.15694/mep.2020.000117.1]
2. Nolan JP, Monsieurs KG, Bossaert L, Böttiger BW, Greif R, Lott C, European Resuscitation Council COVID-Guideline Writing Groups. European Resuscitation Council COVID-19 guidelines executive summary. Resuscitation 2020 Aug;153:45-55 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.06.001] [Medline: 32525022]
3. European Resuscitation Council COVID-19 Guidelines. European Resuscitation Council. 2020 Apr 24. URL: https://www.spci.pt/media/covid-19/European_Resuscitation_Council_COVID-19_Guidelines.pdf [accessed 2020-05-17]
4. Scquizzato T, Olasveengen TM, Ristagno G, Semeraro F. The other side of novel coronavirus outbreak: Fear of performing cardiopulmonary resuscitation. Resuscitation 2020 May;150:92-93 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.03.019] [Medline: 32259607]
5. García-Suárez M, Méndez-Martínez C, Martínez-Isasi S, Gómez-Salgado J, Fernández-García D. Basic life support training methods for health science students: A systematic review. Int J Environ Res Public Health 2019 Mar 03;16(5):768 [FREE Full text] [doi: 10.3390/ijerph16050768] [Medline: 30832440]
6. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. Comput Methods Programs Biomed 2015 Nov;122(2):282-291. [doi: 10.1016/j.cmpb.2015.08.006] [Medline: 26319184]
7. Tobase L, Peres HHC, Gianotto-Oliveira R, Smith N, Polastri TF, Timerman S. The effects of an online basic life support course on undergraduate nursing students' learning. Int J Med Educ 2017 Aug 25;8:309-313 [FREE Full text] [doi: 10.5116/ijme.5985.cbce] [Medline: 28850944]
8. Napp A, Kosan J, Hoffend C, Häge A, Breitfeld P, Doehn C, et al. Implementation of basic life support training for school children: Online education for potential instructors? Results of a cluster randomised, controlled, non-inferiority trial. Resuscitation 2020 Jul;152:141-148. [doi: 10.1016/j.resuscitation.2020.04.041] [Medline: 32422244]

XSL•FO
RenderX

9. Pedersen TH, Kasper N, Roman H, Egloff M, Marx D, Abegglen S, et al. Self-learning basic life support: A randomised controlled trial on learning conditions. Resuscitation 2018 May;126:147-153. [doi: 10.1016/j.resuscitation.2018.02.031] [Medline: 29522830]
10. Finn JC, Bhanji F, Lockey A, Monsieurs K, Frengley R, Iwami T, Education, Implementation, Teams Chapter Collaborators. Part 8: Education, implementation, and teams: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. Resuscitation 2015 Oct;95:e203-e224. [doi: 10.1016/j.resuscitation.2015.07.046] [Medline: 26477705]
11. Rozporządzenie Ministra Edukacji Narodowej z dnia 11 marca 2020 r. w sprawie czasowego ograniczenia funkcjonowania jednostek systemu oświaty w związku z zapobieganiem, przeciwdziałaniem i zwalczaniem COVID-19. Strona główna Sejmu Rzeczpospolitej Polskiej. URL: https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20200000410/O/D20200410.pdf [accessed 2021-05-07]
12. Coyne E, Frommolt V, Rands H, Kain V, Mitchell M. Simulation videos presented in a blended learning platform to improve Australian nursing students' knowledge of family assessment. Nurse Educ Today 2018 Jul;66:96-102. [doi: 10.1016/j.nedt.2018.04.012] [Medline: 29689461]
13. Hu YJ, Ku TH, Jan RH, Wang K, Tseng YC, Yang SF. Decision tree-based learning to predict patient controlled analgesia consumption and readjustment. BMC Med Inform Decis Mak 2012 Nov 14;12:131 [FREE Full text] [doi: 10.1186/1472-6947-12-131] [Medline: 23148492]
14. Staying Alive - AED locations. Staying Alive. URL: https://www.stayingalive.org/index.php?lang=en [accessed 2020-06-20]
15. Gurbanov E. The challenge of grading in self and peer-assessment (undergraduate students' and university teachers' perspectives). Journal of Education in Black Sea Region 2016 May 25;1(2):82-91 [FREE Full text] [doi: 10.31578/jebs.v1i2.21]
16. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance testing: a review of published validity evidence. Med Educ 2011 Apr;45(4):329-338. [doi: 10.1111/j.1365-2923.2010.03863.x] [Medline: 21401680]
17. Naylor K, Torres K. Approaches to stimulate clinical reasoning in continuing medical education during the coronavirus disease 2019 pandemic. Kardiol Pol 2020 Aug 25;78(7-8):770-772 [FREE Full text] [doi: 10.33963/KP.15419] [Medline: 32500994]
18. Spruijt A, Leppink J, Wolfhagen I, Scherpbier A, van Beukelen P, Jaarsma D. Investigating teaching performance in seminars; a questionnaire study with a multi-level approach. BMC Med Educ 2014 Sep 24;14:203 [FREE Full text] [doi: 10.1186/1472-6920-14-203] [Medline: 25253047]
19. Fawns T, Aitken G, Jones D. Online Learning as Embodied, Socially Meaningful Experience. Postdigital Science and Education 2019 May 31;1:293-297 [FREE Full text] [doi: 10.1007/s42438-019-00048-9]
20. Profesjonalny program do organizowania szkoleń BHP online dla firm bhp i pracowników służby bhp!. Szkolenia BHP. URL: https://www.szkolenia-bhp24.pl/ [accessed 2021-05-07]
21. Centrum Medycznych Egzaminów Testowych. Centrum Medycznych Egzaminów Testowych. URL: https://e-csm.umlub.pl/cmet/ [accessed 2021-05-07]
22. Biggs J. Enhancing teaching through constructive alignment. High Educ 1996 Oct;32:347-364. [doi: 10.1007/bf00138871]
23. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. High Educ (Dordr) 2006 Jan 24;31(2):199-218. [doi: 10.1080/03075070600572090]
24. American Heart Association. Heartsaver CPR AED Student Workbook. Dallas, Texas: American Heart Association; 2015.
25. Greif R, Lockey AS, Conaghan P, Lippert A, De Vries W, Monsieurs KG, Education and implementation of resuscitation section Collaborators; Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 10. Education and implementation of resuscitation. Resuscitation 2015 Oct;95:288-301. [doi: 10.1016/j.resuscitation.2015.07.032] [Medline: 26477418]
26. Martínez REB, Ramírez NC, Mesa HGA, Suárez IR, Trejo MDCG, León PP, et al. Decision trees as a tool in the medical diagnosis. Revista Médica de la Universidad Veracruzana 2009;9(2):19-24 [FREE Full text]

Abbreviations

AED: automated external defibrillator
AHA: American Heart Association
BLS: basic life support
CMET: Centrum Medycznych Egzaminów Testowych
ERC: European Resuscitation Council
ILCOR: International Liaison Committee on Resuscitation
OHCA: out-of-hospital cardiac arrest

XSL•FO
RenderX

Edited by C Basch; submitted 11.01.21; peer-reviewed by EJ Faleiros, R Subramaniyam, K Goniewicz, V Kolade; comments to author 27.01.21; revised version received 28.02.21; accepted 22.03.21; published 25.05.21.

Please cite as:
Naylor K, Torres K
Transitioning to Web-Based Learning in Basic Life Support Training During the COVID-19 Pandemic to Battle the Fear of Out-of-Hospital Cardiac Arrest: Presentation of Novel Methods
J Med Internet Res 2021;23(5):e27108
URL: https://www.jmir.org/2021/5/e27108
doi:10.2196/27108
PMID:33886488

©Katarzyna Naylor, Kamil Torres. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Influences of Medical Crowdfunding Website Design Features on Trust and Intention to Donate: Controlled Laboratory Experiment

Xing Zhang[1], PhD; Wenli Hu[1], MSc; Quan Xiao[2], PhD

[1]School of Management, Wuhan Textile University, Wuhan, China

[2]School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China

Corresponding Author:
Quan Xiao, PhD
School of Information Management
Jiangxi University of Finance and Economics
Yuping Avenue
Changbei Economic and Technological Development Zone
Nanchang, 330032
China
Phone: 86 13970840123
Email: xiaoquan@foxmail.com

Abstract

Background: As a type of donation-based crowdfunding, medical crowdfunding has gradually become an important way for patients who have difficulty paying medical bills to seek help from the public. However, many people still have limited confidence in donating money to medical crowdfunding projects.

Objective: Given that the features of a medical crowdfunding website may be important to gain users' trust, this study draws upon two-factor and trust theories to explore how different design features of medical crowdfunding websites affect potential donors' cognition-based trust and affect-based trust, and how these types of trust affect the intention to donate.

Methods: A 2 (informativeness: high vs low) × 2 (visual cues: cool color vs warm color) × 2 (social cues: with vs without) between-subject laboratory experiment was performed to validate our research model. A total of 320 undergraduate students recruited from a university in China participated in the controlled laboratory experiment.

Results: Cognition-based trust (β=.528, P<.001) and affect-based trust (β=.344, P<.001) exerted significant effects on the intention to donate of potential donors of medical crowdfunding. Informativeness as a hygiene factor positively influenced potential donors' cognition-based trust ($F_{1,311}$=49.764, P<.001) and affect-based trust ($F_{1,311}$=16.093, P<.001), whereas social cues as a motivating factor significantly influenced potential donors' cognition-based trust ($F_{1,311}$=38.160, P<.001) and affect-based trust ($F_{1,311}$=23.265, P<.001). However, the color of the webpages affected the two dimensions of trust differently. Specifically, medical crowdfunding webpages with warm colors were more likely to induce affect-based trust than those with cool colors ($F_{1,311}$=17.120, P<.001), whereas no significant difference was found between the effects of cool and warm colors on cognition-based trust ($F_{1,311}$=1.707, P=.19).

Conclusions: This study deepens our understanding of the relationships among the design features of medical crowdfunding websites, trust, and intention to donate, and provides guidelines for managers of medical crowdfunding platforms to enhance potential donors' trust-building by improving the website design features.

(J Med Internet Res 2021;23(5):e25554) doi:10.2196/25554

KEYWORDS

medical crowdfunding; website design; cognition-based trust; affect-based trust; intention to donate

XSL•FO
RenderX

Introduction

Background

Medical expenses are forcing a staggering number of people into poverty worldwide [1]. According to data from the US Census Bureau, 11.2 million Americans are living below the poverty line because of health care costs [2]. Although public health insurance provides coverage for almost all of China's 1.4 billion people, many patients and their families still have difficulty paying medical bills that outpace government insurance provision [3]. Some of them have had to take out loans to pay for their medical expenses, adding to their financial woes and imposing a growing burden of consumer debt on China [4]. As a result, medical crowdfunding, a convenient and effective way to raise funds to alleviate the burden of medical expenses, has skyrocketed in the last few years [5]. For example, GoFundMe, a world-leading crowdfunding platform, reports that more than one-third of the fundraising campaigns it hosts are for health care expenses and that these campaigns raise more than US $650 million annually [6].

In the context of charitable giving, the information asymmetry that exists between charitable project promoters and potential donors is a barrier to successfully soliciting individual donations. Potential donors are often deprived of complete (and updated) information about how their donations will be used [7]. They need charitable project promoters to disclose information about their past performance and governance, as well as information about beneficiaries [8]. To adequately address the information asymmetry problem, and thus increase donations, promoters need to share customized information with targeted donors through appropriate information channels. Medical crowdfunding platforms help patients to make personalized charitable appeals, and to show their illnesses and financial difficulties on their own webpages, which can partially reduce the level of information asymmetry between the crowdfunders and the donors. As such, medical crowdfunding is becoming increasingly popular as it harnesses the power of the crowd to raise funds and reduce the risk of "medical bankruptcy" [5].

Despite the popularity of medical crowdfunding, the failure rate of crowdfunded projects is reported to be high. Many promising projects have failed to meet their fundraising goals for different reasons. The lack of experience in the use of information technology (eg, computers and the internet), insufficient health care literacy, and poor writing skills can all contribute to the failure of a medical crowdfunding project [9]. Fraudulent campaigns, loss of privacy, and small personal social networks may be additional barriers to receive donations from others [9,10]. Hence, many studies have focused on identifying the factors that influence the success of charitable crowdfunding, such as donors' intrinsic and extrinsic motivation [11,12], project content [13], project goal [14], fundraisers' characteristics [13-15], and social influence [16].

Recently, several studies have recognized trust as a critical determinant of funding intention in different crowdfunding contexts [17-20]. In the context of medical crowdfunding, many people have limited confidence in contributing to crowdfunding projects because they are afraid of fraud or do not know whose needs are more urgent. These trust issues result in many medical crowdfunding projects being not fully funded. Therefore, focusing on how to increase people's trust in medical crowdfunding platforms and projects is important. Medical crowdfunding is similar to crowdfunding based on rewards or equity; however, it also has unique characteristics. Medical crowdfunding can essentially be considered as an act of philanthropy, and its success requires not only rationally convincing potential donors but also emotionally motivating them. However, the existing literature on other types of crowdfunding have primarily focused on the effect of a single rational dimension of trust, cognition-based trust [20,21], while largely neglecting the effect of the emotional dimension of trust, affect-based trust. Previous research has noted that emotional appeals are strongly associated with prosocial behavior [22]. Therefore, the first aim of this study was to investigate the role of affect-based trust in a medical crowdfunding context and to examine how such trust affects the intention to donate simultaneously with cognition-based trust.

Despite the increasing number of studies on the relationship between trust and crowdfunding investments, less attention has been paid to the factors that build trust in the context of medical crowdfunding. Therefore, the second focus of this study was on the antecedents of cognition-based and affect-based trust in medical crowdfunding, which can help us to better understand the development of trust. Among the limited research on the antecedents of trust in the field of crowdfunding and charitable giving, most studies have focused on the project type, project content, and fundraisers' expertise and reputation [18-20], whereas few studies have considered how website design features affect potential donors' trust in crowdfunding platforms and projects. Medical crowdfunding websites play an especially important role in publicizing projects and persuading people to donate. On medical crowdfunding websites, fundraisers can write their own charitable appeals to describe their personal experiences of illness and financial difficulties to attract potential donors. As such, medical crowdfunding platforms need to devote considerable efforts to their website design to increase trust and facilitate online donations. Thus, research on design features of medical crowdfunding websites is needed to understand the role of the medical crowdfunding website in influencing people's trust and intention to donate.

To bridge the aforementioned research gaps, in this study, we adopted two-factor and trust theories to identify hygiene and motivation factors in the website design of a medical crowdfunding website, and examined how they affect people's cognition-based trust and affect-based trust, and their ultimate intention to donate.

Trust

Trust in Medical Crowdfunding

The rapid growth of crowdfunding platforms has attracted the attention of scholars, and numerous studies have been performed on the factors that influence crowdfunding success. In these studies, trust has emerged as a key determinant of funding intention in different crowdfunding contexts [17-20]. Trust can be defined as a willingness to accept vulnerability based on beliefs about the trustee's ability and character, and the

XSL•FO
RenderX

emotional bond between the trustor and the trustee [23]. This definition indicates that trust includes not only the intellectual/cognitive dimension (ie, cognition-based trust) but also the emotional/affective dimension (ie, affect-based trust [24]). Table 1 presents a sampling of quantitative research on the role of trust in crowdfunding, revealing the relatively minimal research that has focused on the context of donation-based crowdfunding. Moreover, these studies have provided only partial insight into the two dimensions of trust in crowdfunding. Most of them have conceptualized trust as a one-dimensional construct and revealed that trust in the crowdfunding platform or project creators has a positive effect on individuals' investment decisions [25]. Some other scholars have primarily addressed donors' cognition-based trust, arguing that the credibility and the trustworthiness of the fundraiser and platform are vital factors that influence investors to fund the project [26]. However, little attention has been devoted to the cognitive and affective dimensions of trust simultaneously in the context of crowdfunding.

Similar to other types of crowdfunding, potential donors of medical crowdfunding need to scrutinize the quality of the projects and judge their credibility before deciding to donate. In addition, medical crowdfunding can be regarded as an act of philanthropy where potential donors are incentivized to donate out of sympathy for specific patients seeking help. Therefore, the emotional dimension of trust is also noteworthy because it can evoke an emotional response and inspire people to donate. Hence, this study attempts to contribute to trust theory and the crowdfunding literature by examining whether cognition-based trust and affect-based trust exert positive effects on intention to donate in the context of medical crowdfunding.

Table 1. Representative previous research on the role of trust in crowdfunding.

Reference	Context	Dependent variable	Trust-related variable
Zhang et al [27]	Donation-based crowdfunding	Actual donation	Platform trust
Behl et al [28]	Donation-based crowdfunding	Operational performance	Cognitive trust, swift trust
Kim et al [29]	Incentive crowdfunding	Willingness to crowdfund	Trust in platform, trust in fundraiser
Kim et al [12]	Incentive crowdfunding	Crowdfunding participation	Trust in platform, trust in fundraiser
Chen et al [25]	Donation-based crowdfunding	Intention to donate	Trust
Wehnert et al [30]	Crowdfunding	Trust	Trust
Strohmaier et al [21]	Reward-based crowdfunding	Pledging intention	Trust in creators, trust in the platform
Liang et al [19]	Reward-based crowdfunding	Investment intention	Funder's trust
Cascino et al [26]	Reward-based crowdfunding	Consumer protection regulation, project funding	Perceived credibility of disclosure
Rodriguez-Ricardo et al [31]	Crowdfunding	Intention to participate in crowdfunding	Trust in crowdfunding
Yang et al [32]	Crowdfunding	Investment intention	Trust
Jones and Moncur [33]	Reward-based crowdfunding	Likelihood to invest	Funder's trust
Xiao [34]	Equity-based crowdfunding	Investment intention	Competence trust, relational trust
Klement and Teubner [35]	Equity-based crowdfunding	Number of investments	Trustworthiness

Cognition-Based Trust, Affect-Based Trust, and Intention to Donate

Generally, cognition-based trust is grounded in careful and rational reasoning [36]; it focuses on the "good reasons" why individuals are trustworthy and calls for rational choices based on credible information about the intentions or capabilities of others [37]. Previous research has found that potential sponsors often depend on their evaluation of a crowdfunding project's credibility when they decide whether or not to donate their money to the cause [15,38]. In the field of medical crowdfunding, if potential donors believe that the crowdfunding platform is reliable and the charitable appeal of the fundraiser is logical and reasonable, then their cognition-based trust will be promoted, thereby increasing their intention to donate. Thus, we propose the following hypothesis:

> *H1: Potential donors' cognition-based trust is positively correlated with intention to donate.*

By contrast, affect-based trust focuses on the "emotional bond" between parties. This type of trust is derived not from one's understanding and reasoning but rather from their instincts and feelings [23,37]. Kang et al [18] provided evidence that investors' trust related to emotions has a positive effect on their willingness to invest in crowdfunding platforms. Liu et al [15] also revealed that individuals will increase their willingness to donate because of their affective state derived from the concerns of the charitable crowdfunding fundraiser's situation. Patients

XSL•FO
RenderX

can use crowdfunding platforms to connect emotionally with potential donors to raise funds for health care [39]. Thus, we propose the following hypothesis:

> *H2: Potential donors' affect-based trust is positively correlated with intention to donate.*

Previous studies have argued that individuals will invest further in a relationship only when their basic expectations about the reliability of the relationship are met [24]. Cognition-based trust influences affect-based trust, particularly in the beginning of relationship formation [23]. In many online contexts, cognition-based trust has been empirically demonstrated to be an antecedent of affect-based trust [23,40]. Kang et al [18] investigated the crowdfunding field and found that a rational assessment of the crowdfunding project allows the investor to decide whether to trust the fundraiser, which can enhance their communication and help establish a strong emotional connection. Xu and Wang [41] also argued that logically coherent and complete stories are a prerequisite for eliciting potential donors' empathy and to inspire their charitable behavior. In the context of medical crowdfunding, the cognition-based trust that potential donors develop from carefully reviewing the information provided by the patients can reduce their feelings of insecurity, which in turn arouse emotional responses and facilitate a higher level of affect-based trust. Hence, we propose the following hypothesis:

> *H3: Potential donors' cognition-based trust is positively correlated with their affect-based trust.*

Website Design Features and Trust

Antecedents of Trust in Crowdfunding

Considering that trust can be easily broken in virtual environments, some studies have been performed on how to establish and maintain trust in various contexts [42]. The building of trust in online communities has been wildly studied. For example, Mpinganjira [43] revealed that information usefulness, community responsiveness, and shared vision have a significant influence on consumers' overall trust in health-related virtual communities. Nadeem et al [44] found that social presence of online brand communities, social presence of others, and social presence of interaction have positive effects on social commerce trust. Given that online communities are groups of people with common interests and shared goals, community members gradually develop trust in the community through long-term social interaction. Unlike online communities, people do not need long-term interactions to build trust in a crowdfunding context, but rather rely on a direct and rapid appraisal of the crowdfunding website and the project. Thus, understanding the antecedents of trust in the context of crowdfunding is critical in improving the success rate of crowdfunding projects [45].

Existing research in the field of crowdfunding has mainly examined three categories of antecedent variables of trust: project-, fundraiser-, and platform-related characteristics [15,18,20]. Table 2 provides examples of representative studies focusing on the antecedents of trust. Only a few studies investigated the platform-related characteristics that affect trust. Among these limited investigations, attention was mainly focused on the overall quality or functional features of the platform. For example, Liu et al [15] revealed that website quality and transaction convenience have positive effects on the perceived credibility of a charitable crowdfunding project. Kang et al [18] found that perceived accreditation, structural assurance, and a third-party seal of a crowdfunding platform exert different effects on investors' trust beliefs. However, little attention has been paid to the question of how website design features affect individuals' trust and intention to donate. Website design factors have been demonstrated to be critical in affecting user response and experience [46]. As people learn the details of funding projects by viewing medical crowdfunding websites, exploring the design features of medical crowdfunding websites may contribute to an improved appreciation of trust and intention to donate.

XSL•FO
RenderX

Table 2. Representative previous research on antecedents of trust in crowdfunding.

Reference	Context	Project-related characteristics	Fundraiser-related characteristics	Platform-related characteristics
Kim et al [29]	Incentive crowdfunding	Network externality, perceived informativeness	Value congruence, social interaction ties	Perceived accreditation, structural assurance, third-party seal
Strohmaier et al [21]	Reward-based crowdfunding	N/A[a]	Perceived monitoring	Perceived platform rules, perceived monitoring, perceived pledging security
Liang et al [19]	Reward-based crowdfunding	Project information quality	Fundraiser's ability/expertise, fundraiser's reputation	N/A
Rodriguez-Ricardo et al [31]	Crowdfunding	N/A	Altruism, internal locus of control	N/A
Yang et al [32]	Crowdfunding	Perceived benefits, perceived risk	Communication, shared value	N/A
Kang et al [18]	Reward-based crowdfunding	Network externality, perceived informativeness	Value congruence, social interaction ties	Perceived accreditation, structural assurance
Zheng et al [17]	Reward-based crowdfunding	N/A	Crowdfunding success experience, investment in others	N/A
Liu et al [15]	Donation-based crowdfunding	Project popularity, project content quality	Initiator reputation	Website quality (navigability, security, visual appeal), transaction convenience

[a]N/A: not applicable.

Website Design Features and Two-Factor Theory

A substantial body of literature in the field of information systems has investigated the relationship between website design features and trust. Several website design features have been identified as antecedent variables of trust in various contexts such as navigation design [47,48], privacy and security [47,49], website quality [50], and ease of use [46,48]. However, these antecedents have been examined in a disjointed manner and lack a holistic view of how website design features influence trust. Moreover, existing research has mainly identified the usability and functional features of websites from a cognitive perspective, with less consideration given to the emotional aspects of website design features [46]. Some researchers have demonstrated that a website that meets most usability guidelines does not necessarily lead to higher user favorability [51]. Other scholars have also argued that users' evaluations of a website are often determined by their overall impression of the website, thereby requiring an examination of the holistic aspects of website design features [46,52]. Medical crowdfunding platforms not only need to cognitively enhance the design of their websites to persuade potential donors but should also harness the affective features of their website to evoke potential donors' emotional response and facilitate their willingness to donate. Therefore, in this study, we adopted the Herzberg two-factor theory as a theoretical framework for integrating the cognitive and affective aspects of website design features that influence users' trust in the context of medical crowdfunding.

The two-factor theory was proposed by Herzberg in 1959 [53], which is also known as the hygiene motivation theory. This theory was initially used to explain factors that lead to satisfaction or dissatisfaction in employees' work environment and has gradually been used in a wide range of different domains [54]. Recently, an increasing number of studies have been performed using two-factor theory to examine website design features, and their effects on user attitudes and behaviors in different types of websites [55-57]. In these studies, website design features are systematically distinguished into two categories: hygiene and motivation factors.

Informativeness

Hygiene factors are a collection of attributes closely related to the basic functionality of a website [58]. As the primary purpose of a medical crowdfunding website is to provide the visitors with information about the crowdfunding project, informativeness was selected as the hygiene factor of a medical crowdfunding website in this study. Informativeness can be defined as the extent to which a website offers rich and useful information to users [59,60]; thus, it reflects a website's ability to provide information to its visitors [61]. Previous research has demonstrated that website informativeness is an important prerequisite for users' attitudes toward the website [62], and that it also plays a crucial role in reducing risk and building trust [63]. In the field of crowdfunding, few studies have explored the relationship between informativeness and trust. For example, Kang et al [18] found that funders' cognition-based trust in crowdfunding projects is enhanced if the crowdfunding website provides funders with complete and timely information about the project through a bulletin board. In medical crowdfunding websites, greater informativeness (eg, charitable appeals with several texts and images) helps donors make rational judgments, which in turn makes it easier for them to develop cognition-based trust in the project and the platform. Therefore, we propose the following hypothesis:

XSL•FO
RenderX

H4a: Informativeness is positively correlated with potential donors' cognition-based trust.

Previous research on shopping websites has revealed that the informativeness of a website positively influences consumers' emotional responses [64]. Majumdar and Bose [13] also demonstrated that charitable crowdfunding fundraisers use detailed information to emphasize their needs and suffering to attract others' attention. On the basis of emotional contagion theory, fundraisers' suffering and sadness presented in their stories may increase potential donors' affective reactions. Similarly, detailed information about a medical crowdfunding project not only gives the potential donor insight into the fundraiser's situation but also serves as a means for the fundraiser to get others to listen and alleviate their own suffering. Adding words and images that describe the patient's suffering can result in easily evoking potential donors' affect response and enhancing their emotional bonding with the fundraiser. At the same time, rich project information allows potential donors to know more about the fundraiser and shorten the perceived psychological distance between them, thereby inspiring a higher level of affect-based trust. Hence, we propose the following hypothesis:

H4b: Informativeness is positively correlated with potential donors' affect-based trust.

Visual Cues

Motivation factors are a collection of attributes that are closely related to the additional value-added services offered by a website [58]. Previous studies have found that in contrast to hygiene factors, motivation factors are crucial for website users' satisfaction [55,57]. In the medical crowdfunding context, motivation factors are the website design features of a medical crowdfunding website that stimulate potential donors to make donation decisions beyond meeting their basic information needs. In this study, the website's visual cues (ie, warm or cool colors) and social cues were selected as motivation factors.

The visual design of a website has proven to be an impactful determinant of trust [48]. Among the many visual design elements, color serves as a specific visual stimulus that shapes a visitor's perception of the ambient temperature [65]. Such perceived cool or warmth on temperature elicits varying degrees of social proximity, language concreteness, and relational focus, and exerts an effect on individual behavior through the mechanism of insula [66]. Williams and Bargh [67] showed that the warmth sensed by the body can influence interpersonal judgments and prosocial behavior. Literature from several fields has noted that red and orange tend to evoke warmth, whereas blue and green are reminiscent of cold [68]; this comparison seems to convey a more positive role of warm colors. However, an ever-increasing number of studies in recent years have challenged this view by emphasizing the effects of cool colors. Rizomyliotis et al [69] found that cool backgrounds are more likely to inspire stronger positive attitudes and behavioral intentions than warm colors, and cool colors usually create a sense of relaxation and calm [70]. On the one hand, cool colors are often linked to competence, trust, safety, and sincerity, such that many companies incorporate blue into their brand logos to create a professional and trustworthy image, and consumers regard blue logos to be more reliable than other colors [71]. On the other hand, in terms of the information processing of individuals, Ettis [72] suggested that cool colors (eg, blue) facilitate individuals' information seeking, processing, and consideration. Therefore, medical crowdfunding sites with more cool-colored elements may be more conducive to donor attention, and could promote rational and logical reasoning about the help-seeking information on the platform, which in turn will enhance the donors' cognition-based trust in the platform and the project. Thus, we propose the following hypothesis:

H5a: Potential donors' cognition-based trust in a medical crowdfunding website with a cool-toned (blue) interface is higher than that in a warm-toned (orange) interface.

Previous studies have documented that perceived warmth is associated with interpersonal intimacy; that is, perceived warmth creates feelings of interpersonal closeness [73], and promotes interpersonal trust, cooperation, and friendship [74]. Specifically, warm colors are more evocative of warmth than cool colors [66] and provide consumers with a sense that their surroundings are more socially dense [75]. In this manner, a warm-toned interface of a medical crowdfunding website will bring warm psychological perceptions and positive attitudes to donors, making them feel less distant from the medical crowdfunding platform and other help-seekers, and thus more likely to evoke their emotional trust toward the platform and project. Hence, we posit the following hypothesis:

H5b: Potential donors' affect-based trust in a medical crowdfunding website with a warm-toned (orange) interface is higher than that in a cool-toned (blue) interface.

Social Cues

Social cues in web design refer to social presence and social interaction embedded in the web interface through various communication tools [76]. These cues have been shown to have important influences on individuals' cognition-based trust in various contexts such as online shopping and e-commerce. For example, early consumer endorsement in the form of online reviews in e-commerce can help other consumers determine the usefulness of a product, build cognition-based trust, and thus facilitate rational shopping decisions [77]. Social cues can serve as a collective endorsement, allowing the crowdfunding fundraisers to signal credibility [78]. In medical crowdfunding, fundraisers need to reach out to others through social networks to attract more donations. Therefore, the presence of social cues (eg, proofs and reviews from other donors) helps to exhibit the authenticity and credibility of a project, thereby inspiring a higher level of cognition-based trust from potential donors. On this basis, we posit the following hypothesis:

H6a: Potential donors' cognition-based trust in a medical crowdfunding website with social cues is higher than that without social cues.

The presentation of social cues on a website leads to a perception of warmth and friendliness, which triggers a sense of social presence [79]. Given that social presence is strongly associated

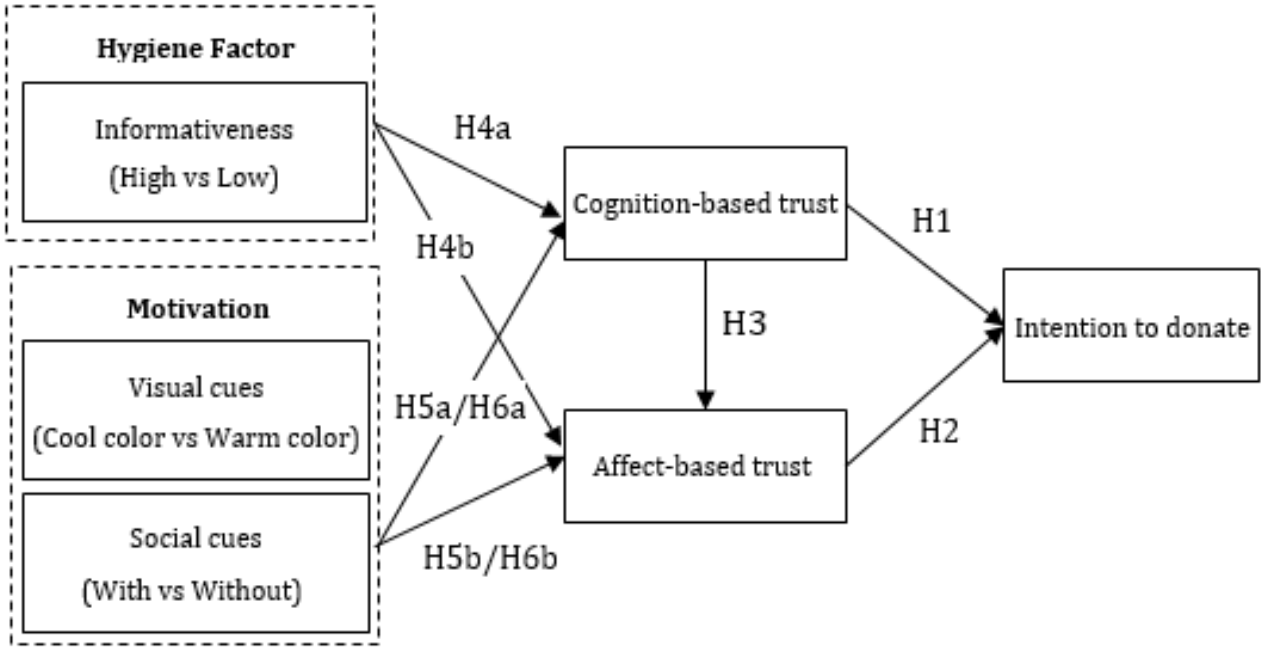
XSL•FO
RenderX

with emotional responses and interpersonal warmth [80,81], the use of social cues will help users build affect-based trust [79]. Furthermore, affect-based trust also indicates that the trustor builds trust in the trustee according to their benevolence and moral norms [82]. In medical crowdfunding websites, social cues such as other donors' feedback and the fundraiser's replies can be used as third-party opinions to evaluate the fundraiser's benevolence and moral norms, which enhance affect-based trust-building. Thus, we posit the following hypothesis:

> *H6b: Potential donors' affect-based trust in a medical crowdfunding website with social cues is higher than that without social cues.*

On the basis of these hypotheses, we propose the research model shown in Figure 1.

Figure 1. Research model.



Methods

Experimental Design

To test how the design features of medical crowdfunding websites influence potential donors' intention to donate, we performed a laboratory experiment with a 2 (informativeness: high vs low) × 2 (visual cues: cool color vs warm color) × 2 (social cues: with vs without) between-subject factorial design to test the hypotheses. We designed eight medical crowdfunding webpages for the experiment.

The webpages were developed to mimic the layout of the mobile versions of Qingsongchou, a major medical crowdfunding platform in China, to enhance the external validity. A fictional medical crowdfunding project was shown in these webpages. The project began with a charity appeal about a patient with cancer who cannot pay his medical bills. We chose cancer for two reasons. First, cancer is one of the most common disease types with a high mortality rate. Second, we crawled the information of medical crowdfunding projects on the Qingsongchou website, and found that cancer is the most common cause in medical crowdfunding projects. To protect patient privacy, we made changes to basic patient information (eg, profile picture, nickname, name, and home address). Eight screenshots of the experimental webpages are presented in Multimedia Appendix 1.

Variable Manipulations

Approach

We used a laboratory experimental approach because it allowed us to examine the effects of website design features on users' intention to donate. Three website design features needed to be manipulated in our study: informativeness, visual cues, and social cues.

Manipulations of Informativeness

This study manipulated the informativeness of the medical crowdfunding webpages into two groups: high informativeness and low informativeness. The high informativeness group presented more categories of information (ie, patient's charity appeal, goals and progress of crowdfunding, medical certificate, proof of the patient's financial hardship, and information about the disease) in the webpage, along with more pictures and longer text. The low informativeness group contained fewer categories of information (ie, only the patient's charity appeal, goals and progress of crowdfunding, and medical certificate), along with fewer pictures and shorter text.

Manipulations of Visual Cues

The visual cues in this study focused on the overall color of the webpages, including the background color of the pages and the font colors. The main color of the pages in the cool color groups was blue (hue: 170, saturation: 239, luminance: 128, transparency: 90%), whereas the main color of the warm color

XSL•FO
RenderX

groups was orange (hue: 15, saturation: 239, luminance: 128, transparency: 90%) [83].

Manipulations of Social Cues

Webpages in the "with social cues" group showed the buttons for retweeting the project to social media, social endorsement from others, and information about other people's donations. Those in the "without social cues" groups did not contain such information.

Measurement

This research model includes three latent constructs: cognition-based trust, affect-based trust, and intention to donate. We selected the scales of these constructs from previous research with adaptive modifications to fit the context of medical crowdfunding. All of the measurements were scored on a 7-point scale (1: absolutely disagree, 7: absolutely agree). To eliminate the influence of the participants' mood on the results of the experiment, we introduced participants' mood as a covariate in the research model. Mood was measured by the item "How would you rate your current emotional state?", which was scored from 1 (negative) to 7 (positive). The scale showed good reliability and validity.

We used the backtranslation method to translate the original English scale into the Chinese version and compared the two versions of the scales to check for translation accuracy. Four experts in the field of information systems were invited to evaluate the content validity of the scales. Following the opinions of the survey experts, we made some revisions to the items, such as revising the wording to improve clarity and accuracy, shortening the length of items to reduce survey fatigue, modifying ambiguous items, and revising items to make them more tailored to the medical crowdfunding context. Table 3 shows the three latent constructs and their measurement items.

Table 3. Research constructs and measurements.

Construct	Measurement items	References
Cognition-based trust	(1) I can rely on this medical crowdfunding platform and project; (2) the medical crowdfunding platform and project have my confidence; (3) the medical crowdfunding platform and project have high integrity	Kang et al [18], Ranganathan et al [84]
Affect-based trust	(1) If I share my problems with this medical crowdfunding platform, it would respond caringly; (2) the medical crowdfunding platform displays a warm and caring attitude toward me; (3) I can talk freely with the medical crowdfunding platform about my problems	Ranganathan et al [84], Johnson and Grayson [85]
Intention to donate	(1) The probability that I would donate money to the medical crowdfunding project is high; (2) my donation intention to the medical crowdfunding project is high; (3) the likelihood of my donating money to the medical crowdfunding project is high	Liu et al [15]

Data Collection

A total of 320 undergraduate students recruited from a university in China participated in the controlled laboratory experiment. Each participant was assigned to one of eight webpages varying in informativeness, color, and social cues. The participants were then asked to complete the questionnaire after viewing the webpage. The age of the participants ranged from 18 to 26 years (mean 21.719 years). Analysis of variance showed no significant differences in gender ($F_{7,312}$=1.097, P=.37) or age ($F_{7,312}$=1.690, P=.11) across the eight experimental conditions. All subjects were randomly assigned to one of these eight treatment groups and completed the questionnaire after viewing their assigned webpages.

Results

Manipulation Checks

We performed manipulation checks using the measurement items adapted from previous studies. The item "The medical crowdfunding platform offers relevant, timely, and accurate information about the project" was adopted from the works of Kang et al [18] and Kim et al [86] to measure informativeness. This item was scored on a 7-point Likert scale (1: absolutely disagree, 7: absolutely agree). The item "To what extent do you feel warm when you see the colors of the website's interface?" was adopted from the work of Motoki et al [87] to measure the warmth of the webpage's color, which was also rated on a 7-point Likert scale (1: not warm at all to 7: very warm). We also checked the presence of social cues by adapting the scale of Friedrich et al [88], "Have you noticed the social cues on the medical crowdfunding site?", with the item rated on a 3-point scale (1=no, 2=not sure, 3=yes).

The results from a t test revealed that subjects assigned to the high informativeness conditions agreed more that the webpage has richer information than those in low informativeness conditions (mean 4.925 vs 3.031, t_{318}=17.618; P<.001). Similarly, the t test results showed that the participants who viewed webpages with orange color felt warmer than those who viewed pages with blue as the dominant color (mean 4.875 vs 2.663, t_{318}=18.829; P<.001). The t test results also suggested that the score for the webpages with social cues was higher than that of pages without social cues (mean 2.838 vs 1.250, t_{318}=29.088; P<.001). All of the results of the manipulation check were significant.

Measurement Validation

Cronbach α for checking the internal consistency was initially computed using SPSS 24.0. Table 4 shows that the Cronbach α of each construct exceeded .90, which is higher than the recommended value of .70. A confirmatory factor analysis was then performed using AMOS 22.0 to examine the convergent validity of the constructs. Table 4 shows that the values of factor loadings and composite reliability of all constructs were greater than 0.70, and the average variance extracted (AVE) exceeded 0.5. These results showed good convergent validity [89].

XSL•FO
RenderX

Table 4. Convergent validity and internal reliability.

Construct	Factor loading	Cronbach α	AVE[a]	CR[b]
Cognition-based trust (CT)		.909	0.773	0.911
CT1	0.814			
CT2	0.907			
CT3	0.913			
Affect-based trust (AT)		.865	0.686	0.867
AT1	0.780			
AT2	0.848			
AT3	0.854			
Intention to donate (ID)		.915	0.784	0.916
ID1	0.881			
ID2	0.903			
ID3	0.872			

[a]AVE: average variance extracted.

[b]CR: construct reliability.

Finally, we computed the square root of the AVE values of all constructs and the correlation coefficients between these constructs to test the discriminant validity. The square roots of the AVE values were greater than other correlation coefficients, which indicated adequate discriminant validity [90].

Hypothesis Testing

We utilized AMOS22.0 to test the hypotheses regarding the effects of cognition-based trust and affect-based trust on intention to donate. As predicted, cognition-based trust had a significant and positive influence on intention to donate (β=.528, *P*<.001), supporting H1. The effect of affect-based trust on intention to donate was significant and positive (β=.344, *P*<.001), supporting H2. In addition, cognition-based trust significantly influenced affect-based trust (β=.641, *P*<.001), supporting H3. Thus, H1-H3 were supported.

Given that we have two dependent variables (ie, cognition-based trust and affect-based trust), we performed multivariate analysis of variance to check whether the experimental treatments had a general effect on both variables. Three website design features (ie, informativeness, visual cues, and social cues) were modeled as the fixed factors, and participants' mood was included as a covariate. The main effects of informativeness (Wilks λ=0.862, F_{310}=24.870; *P*<.001), visual cues (Wilks λ=0.902, F_{310}=16.846; *P*<.001), and social cues (Wilks λ=0.882, F_{310}=20.741; *P*<.001) were significant. We then performed two analyses of covariance (ANCOVAs) on cognition-based trust and affect-based trust separately.

A three-way ANCOVA was initially performed on cognition-based trust. The main effect of informativeness on cognition-based trust was significant ($F_{1,311}$=49.764, *P*<.001), indicating that higher informativeness is associated with higher donor cognition-based trust in the platform and the project. Thus, H4a was supported. Similarly, social cues had a significant effect on cognition-based trust ($F_{1,311}$=38.160, *P*<.001). In other words, donors' cognition-based trust is higher when social cues are present in medical crowdfunding webpages. Thus, H6a was supported. However, visual cues had no significant effect on cognition-based trust ($F_{1,311}$=1.707, *P*=.19). Thus, H5a was not supported. Moreover, no significant interaction effect was observed among informativeness, visual cues, and social cues.

A three-way ANCOVA was then performed on affect-based trust. The results indicated a significant main effect of informativeness ($F_{1,311}$=16.093, *P*<.001), visual cues ($F_{1,311}$=17.120, *P*<.001), and social cues ($F_{1,311}$=23.265, *P*<.001). Thus, H4b, H5b, and H6b were supported. Moreover, no interaction effect was found.

Mediation Analysis

To further verify whether the two different types of trust mediate the effects of website design features on intention to donate, we performed a mediation analysis in SPSS 24.0 with the PROCESS plugin developed by Hayes [91]. First, we used cognition-based trust as the independent variable, affect-based trust as the mediator, and intention to donate as the dependent variable. A mediation bootstrapping analysis with 5000 bootstrap samples revealed that the bootstrap 95% CI did not include 0 (β=.185, 95% CI 0.120-0.261), showing that affect-based trust mediates the effect of cognition-based trust on intention to donate.

Next, we used intention to donate as the dependent variable, three website design features (ie, informativeness, social cues, and visual cues) as independent variables, and cognition-based trust as the mediator to test the mediating mechanisms involved. The effects of informativeness (β=.545, 95% CI 0.380-0.721) and social cues (β=.454, 95% CI 0.297-0.629) on intention to donate were mediated by cognition-based trust. However, cognition-based trust did not mediate the relationship between visual cues and intention to donate (β=−.102, 95% CI –0.271 to 0.692).

XSL•FO
RenderX

Finally, we examined the mediation role of affect-based trust. Affect-based trust mediated the relationship between informativeness (β=.159, 95% CI 0.105-0.214), visual cues (β=.325, 95% CI 0.178-0.482), social cues (β=.299, 95% CI 0.141-0.464), and intention to donate.

Discussion

Principal Findings

This study reveals that cognition- and affect-based trust significantly influence medical crowdfunding users' intention to donate. This finding is in line with previous studies that demonstrate the salient impact relationship between trust and intention to donate [17,19]. However, the difference is that this study did not distinguish among the different dimensions of trust [19,20,92] and instead only examined cognition-based trust [17]. This study validates the effects of different dimensions of trust on intention to donate in the context of medical crowdfunding. The results show that not only does cognition-based trust have a significant effect on intention to donate for medical crowdfunding projects but so does affect-based trust. We further found that affect-based trust plays a mediating role in the relationship between cognition-based trust and intention to donate.

Furthermore, this study introduced Hertzberg two-factor theory into the research model and identified two types of medical crowdfunding website design features (ie, hygiene and motivator factors). The results indicate that informativeness as a hygiene factor positively influences potential donors' cognition-based trust and affect-based trust. Among the motivation factors, social cues significantly influence potential donors' cognition-based trust and affect-based trust. However, webpage color as a visual cue affects the two dimensions of trust differently. Specifically, medical crowdfunding webpages with a warm color were more likely to induce affect-based trust than those with a cool color, whereas no significant difference was found between the effects of cool and warm colors on cognition-based trust. One possible explanation for this finding is that cognitive processes are not instantaneous, and cognition-based trust is formed only after individuals can cognitively process and evaluate available evidence [93,94]. The evidence presented on medical crowdfunding websites has a bearing on whether a patient can be cured of a serious illness and escape bankruptcy due to the illness, which requires the potential donors to devote energy to evaluate and extrapolate. Although different hues of color can cause potential donors to perceive temperature and interpersonal warmth differently, these perceptions are not directly responsible for the differences in their cognition-based trust.

Our results also indicate that cognition- and affect-based trust have mediating effects on the relationship between website design features and intention to donate. Different website design features have different influences on cognition- and affect-based trust, which in turn exert different effects on intention to donate. Particularly, informativeness and social cues indirectly affect intention to donate through cognition- and affect-based trust. Webpage colors significantly affected intention to donate through affect-based trust, whereas cognition-based trust had no mediating effect on the relationship between webpage colors and intention to donate.

Implications

Theoretical Implications

The theoretical contributions of this study are as follows. First, this study contributes to the theory of trust and enriches the literature on crowdfunding. Despite the existence of impressive studies that have examined trust and willingness to fund in different crowdfunding contexts, few of them have discussed trust issues in the context of medical crowdfunding. This study identifies the importance of affect-based trust in medical crowdfunding, which is often ignored in other crowdfunding contexts. Specifically, affect-based trust not only influences intention to donate for medical crowdfunding projects simultaneously with cognition-based trust but it also mediates the influence of cognition-based trust on intention to donate.

Second, identifying the antecedents that influence cognition- and affect-based trust is important for trust development. This study explored the holistic aspects of the design features of medical crowdfunding websites to understand how users develop trust to medical crowdfunding platforms and projects. We applied Hertzberg two-factor theory to identify hygiene and motivation factors in website design features as antecedent variables that influence the two types of trust. We validated the positive influence of informativeness as a hygiene factor of website design on cognition- and affect-based trust, and demonstrated the motivating effects of visual and social cues on trust. These findings enrich the literature on website design by broadening the application of two-factor theory to the design of medical crowdfunding websites.

Third, previous studies have briefly introduced the relationships between website design features and users' behavioral intention; however, few studies have attempted to reveal the mechanisms underlying the effects of website design features on the intention to donate in medical crowdfunding. By integrating the design features of medical crowdfunding websites, trust, and intention to donate, we contribute to filling this gap, while demonstrating the mediating role of cognition-based trust and affect-based trust on such impacts.

Practical Implications

This study has important implications for medical crowdfunding platform operators and users. First, our empirical results reveal the significant effects of cognition- and affect-based trust on intention to donate. Therefore, platform operators and project initiators of medical crowdfunding should not only enhance the authenticity and credibility of their platforms and projects but also emphasize their affective elements.

Second, the informativeness of the projects presented on medical crowdfunding platforms is a hygiene factor that can elicit users' trust. Medical crowdfunding platform operators should develop more functions that contribute to the information richness of the website, such as providing a wider range of information categories and the functions of uploading pictures and videos. Moreover, project initiators need to elaborate more on their

XSL•FO
RenderX

situations and charitable appeals to inspire potential donors logically and emotionally.

Third, visual and social cues on medical crowdfunding platforms also have significant effects on potential donors' trust. This finding provides medical crowdfunding website designers with some guidelines. They can enhance the usage of warm colors to create a warm and caring atmosphere to boost users' affect-based trust. They can also provide more social cues such as other peoples' comments to increase potential donors' cognition- and affect-based trust.

Limitations and Future Research Directions

Despite the contributions of this study, we have identified several limitations. First, the subjects in this experiment were undergraduate students. Our primary reason for choosing students as subjects was that students are easier to recruit than other social groups and are more obedient, which is beneficial to the success of laboratory experiments. In addition, as Chinese medical crowdfunding projects are mainly diffused through internet apps such as Weibo and WeChat, social media users have become the main source of donations. Weibo users under the age of 25 account for 57.4% of the total number of users [95]. In terms of age distribution, students are an important user group for social media and are thus the main potential donors for medical crowdfunding. However, despite the representativeness of students, they still do not accurately represent the overall population of medical crowdfunding users. Therefore, future research will need to recruit a larger population to improve the representativeness of the subjects.

Second, this study focused on three website design features based on the previous literature and the context of medical crowdfunding. However, there may be other website design features that can influence users' trust and donations. For example, we only focused on color in visual cues, whereas font styles or page layout may also affect users' trust in the platform and the project. Future research can identify more website design features of relevance.

Third, this study manipulated informativeness as a one-dimensional variable incorporating text length, number of pictures, and categories of information without considering the interactions among these features. Future research can introduce new content-related features and further explore the interactions among these factors.

Finally, individuals may perceive colors in varied and nuanced ways. In our study, we only focused on one color feature that is most likely to be noticed by the user (ie, hue) and ignored the other features of color. Future research can explore the effects of other color features (eg, saturation and luminance) on potential donors' trust.

Conclusions

The purpose of this study was to explore the role of cognition- and affect-based trust in the medical crowdfunding context and to test the effects of medical crowdfunding website design features on these two dimensions of trust. By applying trust theory and Hertzberg two-factor theory, we identified several hygiene and motivation website design factors, and confirmed their influencing mechanisms on trust in our empirical study. This study not only enriches the literature on crowdfunding but also provides implications for medical crowdfunding platform operators and users on how to promote trust.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (71974152 and 71861014). We deeply appreciate the Research Center of Enterprise Decision Support, a key research institute of humanities and social sciences in the universities of Hubei province.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eight screenshots of the experimental webpages.
[RAR File , 12683 KB - jmir_v23i5e25554_app1.rar]

References

1. Aizenman N. Health care costs push a staggering number of people into extreme poverty. National Public Radio. 2017 Dec 14. URL: https://www.npr.org/sections/goatsandsoda/2017/12/14/569893722/health-care-costs-push-a-staggering-number-of-people-into-extreme-poverty [accessed 2020-04-25]
2. Konard W. Health care costs still push Americans into poverty. CBS News. 2016 Sep 13. URL: https://www.cbsnews.com/news/health-care-costs-still-push-americans-into-poverty/ [accessed 2021-02-05]
3. Huang ZL, Shan J. China aims at poverty caused by medical bills. China Daily. 2017 Jun 09. URL: http://www.chinadaily.com.cn/china/2017-06/09/content_29680299.htm [accessed 2021-02-05]
4. Reuters. China healthcare costs forcing patients into crippling debt. Fortune. 2016 Jul. URL: https://fortune.com/2016/07/10/china-healthcare-costs-debt/ [accessed 2020-04-23]
5. Burtch G, Chan J. Investigating the relationship between medical crowdfunding and personal bankruptcy in the United States: evidence of a digital divide. MIS Quart 2019 Jan 1;43(1):237-262. [doi: 10.25300/misq/2019/14569]

XSL•FO
RenderX

6. Hunt J. Crowdfunding strategies for medical expenses. The Balance. 2019 Mar. URL: https://www.thebalance.com/crowdfunding-strategies-for-medical-expenses-4184814 [accessed 2020-04-25]
7. Beldad A, Gosselt J, Hegner S, Leushuis R. Generous but not morally obliged? Determinants of Dutch and American Donors' Repeat Donation Intention (REPDON). Voluntas 2014 Jan 18;26(2):442-465. [doi: 10.1007/s11266-013-9419-9]
8. Ebrahim A. The many faces of nonprofit accountability. In: Renz DO, Herman RD, editors. The Jossey-Bass handbook of nonprofit leadership and management. 3rd edition. San Francisco: Jossey-Bass; 2010:101.
9. Heath S. Does medical crowdfunding perpetuate affordability problems? Patient Engagement HIT. 2020 Oct 28. URL: https://patientengagementhit.com/news/does-medical-crowdfunding-perpetuate-affordability-problems [accessed 2021-02-06]
10. Berliner LS, Kenworthy NJ. Producing a worthy illness: Personal crowdfunding amidst financial crisis. Soc Sci Med 2017 Aug;187:233-242. [doi: 10.1016/j.socscimed.2017.02.008] [Medline: 28274601]
11. Gleasure R, Feller J. Does heart or head rule donor behaviors in charitable crowdfunding markets? Int J E Commerce 2016 Jun 20;20(4):499-524. [doi: 10.1080/10864415.2016.1171975]
12. Kim MJ, Bonn M, Lee C. The effects of motivation, deterrents, trust, and risk on tourism crowdfunding behavior. Asia Pacif J Tourism Res 2019 Nov 12;25(3):244-260. [doi: 10.1080/10941665.2019.1687533]
13. Majumdar A, Bose I. My words for your pizza: An analysis of persuasive narratives in online crowdfunding. Inf Manage 2018 Sep;55(6):781-794. [doi: 10.1016/j.im.2018.03.007]
14. Mollick E. The dynamics of crowdfunding: An exploratory study. J Bus Ventur 2014 Jan;29(1):1-16. [doi: 10.1016/j.jbusvent.2013.06.005]
15. Liu L, Suh A, Wagner C. Empathy or perceived credibility? An empirical study on individual donation behavior in charitable crowdfunding. Internet Res 2018 Jun 04;28(3):623-651. [doi: 10.1108/intr-06-2017-0240]
16. Agrawal A, Catalini C, Goldfarb A. Crowdfunding: geography, social networks, and the timing of investment decisions. J Econ Manage Strateg 2015 May 04;24(2):253-274. [doi: 10.1111/jems.12093]
17. Zheng H, Hung J, Qi Z, Xu B. The role of trust management in reward-based crowdfunding. Online Inf Rev 2016 Feb 08;40(1):97-118 [FREE Full text] [doi: 10.1108/OIR-04-2015-0099]
18. Kang M, Gao Y, Wang T, Zheng H. Understanding the determinants of funders' investment intentions on crowdfunding platforms. Indust Manag Data Syst 2016 Sep 12;116(8):1800-1819. [doi: 10.1108/imds-07-2015-0312]
19. Liang T, Wu SP, Huang C. Why funders invest in crowdfunding projects: Role of trust from the dual-process perspective. Inf Manage 2019 Jan;56(1):70-84. [doi: 10.1016/j.im.2018.07.002]
20. Moysidou K, Hausberg JP. In crowdfunding we trust: A trust-building model in lending crowdfunding. J Small Bus Manag 2019 Nov 08;58(3):511-543. [doi: 10.1080/00472778.2019.1661682]
21. Strohmaier D, Zeng J, Hafeez M. Trust, distrust, and crowdfunding: A study on perceptions of institutional mechanisms. Telemat Informat 2019 Oct;43:101252. [doi: 10.1016/j.tele.2019.101252]
22. Kemp E, Kennett-Hensel PA, Kees J. Pulling on the heartstrings: examining the effects of emotions and gender in persuasive appeals. J Advertising 2013 Jan;42(1):69-79. [doi: 10.1080/00913367.2012.749084]
23. Wang W, Qiu L, Kim D, Benbasat I. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decis Support Syst 2016 Jun;86:48-60. [doi: 10.1016/j.dss.2016.03.007]
24. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 1995 Feb 01;38(1):24-59. [doi: 10.2307/256727]
25. Chen Y, Dai R, Yao J, Li Y. Donate time or money? The determinants of donation intention in online crowdfunding. Sustainability 2019 Aug 07;11(16):4269. [doi: 10.3390/su11164269]
26. Cascino S, Correia M, Tamayo A. Does consumer protection enhance disclosure credibility in reward crowdfunding? J Account Res 2019 Nov 15;57(5):1247-1302. [doi: 10.1111/1475-679x.12289]
27. Zhang Y, Tan CD, Sun J, Yang Z. Why do people patronize donation-based crowdfunding platforms? An activity perspective of critical success factors. Comput Hum Behav 2020 Nov;112:106470. [doi: 10.1016/j.chb.2020.106470]
28. Behl A, Dutta P, Sheorey P, Singh RK. Examining the role of dialogic communication and trust in donation-based crowdfunding tasks using information quality perspective. TQM J 2020 Aug 19:online ahead of print. [doi: 10.1108/tqm-06-2020-0139]
29. Kim MJ, Hall CM, Kim D. Why do investors participate in tourism incentive crowdfunding? The effects of attribution and trust on willingness to fund. J Travel Tourism Market 2020 Feb 12;37(2):141-154. [doi: 10.1080/10548408.2020.1722784]
30. Wehnert P, Baccarella CV, Beckmann M. In crowdfunding we trust? Investigating crowdfunding success as a signal for enhancing trust in sustainable product features. Technol Forecast Soc Change 2019 Apr;141:128-137. [doi: 10.1016/j.techfore.2018.06.036]
31. Rodriguez-Ricardo Y, Sicilia M, López M. Altruism and internal locus of control as determinants of the intention to participate in crowdfunding: the mediating role of trust. J Theor Appl Electron Commer Res 2019 Sep;14(3):1-16. [doi: 10.4067/s0718-18762019000300102]
32. Yang X, Zhao K, Tao X, Shiu E. Developing and validating a theory-based model of crowdfunding investment intention—perspectives from social exchange theory and customer value perspective. Sustainability 2019 May 01;11(9):2525. [doi: 10.3390/su11092525]

XSL•FO
RenderX

33. Jones H, Moncur W. A mixed-methods approach to understanding funder trust and due diligence processes in online crowdfunding investment. Trans Soc Comput 2020 Feb 28;3(1):1-29. [doi: 10.1145/3373148]
34. Xiao L. How lead investors build trust in the specific context of a campaign. Int J Entrepren Behav Res 2019 Dec 06;26(2):203-223. [doi: 10.1108/IJEBR-05-2019-0265]
35. Klement F, Teubner T. Trust isn't blind: exploring visual investor cues in equity crowdfunding. 2019 Dec Presented at: International Conference on Information Systems (ICIS) 2019; December 15, 2019; Munich, Germany p. 1-9.
36. Lewis JD, Weigert A. Trust as a social reality. Soc Forces 1985 Jun 01;63(4):967-985. [doi: 10.1093/sf/63.4.967]
37. Punyatoya P. Effects of cognitive and affective trust on online customer behavior. Market Intell Plan 2019 Feb 04;37(1):80-96. [doi: 10.1108/MIP-02-2018-0058]
38. Greenberg M, Gerber E. Learning to fail: experiencing public failure online through crowdfunding. New York: Association for Computing Machinery; 2014 Apr Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April 2014; Toronto, Ontario, Canada p. 581-590. [doi: 10.1145/2556288.2557110]
39. Moore B. Medical crowdfunding and the virtuous donor. Bioethics 2019 Feb;33(2):238-244. [doi: 10.1111/bioe.12527] [Medline: 30341924]
40. Kanawattanachai P, Yoo Y. Dynamic nature of trust in virtual teams. J Strat Inf Syst 2002 Dec;11(3-4):187-213. [doi: 10.1016/s0963-8687(02)00019-7]
41. Xu K, Wang X. "Kindhearted people, please save my family": narrative strategies for new media medical crowdfunding. Health Commun 2020 Nov;35(13):1605-1613. [doi: 10.1080/10410236.2019.1654173] [Medline: 31455114]
42. Hsu M, Chang C, Yen C. Exploring the antecedents of trust in virtual communities. Behav Inf Technol 2011 Sep;30(5):587-601. [doi: 10.1080/0144929x.2010.549513]
43. Mpinganjira M. Precursors of trust in virtual health communities: A hierarchical investigation. Inf Manag 2018 Sep;55(6):686-694. [doi: 10.1016/j.im.2018.02.001]
44. Nadeem W, Khani AH, Schultz CD, Adam NA, Attar RW, Hajli N. How social presence drives commitment and loyalty with online brand communities? the role of social commerce trust. J Retail Consum Serv 2020 Jul;55:102136. [doi: 10.1016/j.jretconser.2020.102136]
45. Shankar V, Urban GL, Sultan F. Online trust: a stakeholder perspective, concepts, implications, and future directions. J Strateg Inf Syst 2002 Dec;11(3-4):325-344. [doi: 10.1016/s0963-8687(02)00022-7]
46. Pengnate S, Sarathy R. An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors. Comput Hum Behav 2017 Feb;67:49-60. [doi: 10.1016/j.chb.2016.10.018]
47. Ganguly B, Dash SB, Cyr D. Website characteristics, trust and purchase intention in online stores:-an empirical study in the Indian context. J Inf Sci Technol 2009 Jan;6(2):22-44.
48. Cyr D. Modeling web site design across cultures: relationships to trust, satisfaction, and e-loyalty. J Manag Inf Syst 2014 Dec 08;24(4):47-72. [doi: 10.2753/MIS0742-1222240402]
49. Frik A, Mittone L. Factors influencing the perception of website privacy trustworthiness and users' purchasing intentions: the behavioral economics perspective. J Theor Appl Electron Commer Res 2019 Sep;14(3):89-125. [doi: 10.4067/s0718-18762019000300107]
50. Faisal CMN, Gonzalez-Rodriguez M, Fernandez-Lanvin D, de Andres-Suarez J. Web design attributes in building user trust, satisfaction, and loyalty for a high uncertainty avoidance culture. IEEE Trans Human-Mach Syst 2017 Dec;47(6):847-859. [doi: 10.1109/thms.2016.2620901]
51. Hart TA, Chaparro BS, Halcomb CG. Evaluating websites for older adults: adherence to 'senior-friendly' guidelines and end-user performance. Behav Inf Technol 2008 May;27(3):191-199. [doi: 10.1080/01449290600802031]
52. Deng L, Poole MS. Affect in web interfaces: a study of the impacts of web page visual complexity and order. MIS Quart 2010 Dec;34(4):711-730. [doi: 10.2307/25750702]
53. Sanjeev MA, Surya AV. Two factor theory of motivation and satisfaction: an empirical verification. Ann Data Sci 2016 Mar 30;3(2):155-173. [doi: 10.1007/s40745-016-0077-9]
54. Wu TL, Deng ZH, Chen Z, Zhang DL, Wang RX, Wu X. Predictors of patients' intention to interact with doctors in web-based health communities in China: cross-sectional study. J Med Internet Res 2019 Jun 07;21(6):e13693 [FREE Full text] [doi: 10.2196/13693] [Medline: 31199296]
55. Zhang P, von Dran GM. Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. J Am Soc Inf Sci 2000 Oct;51(14):1253-1268. [doi: 10.1002/1097-4571(2000)9999:9999<::aid-asi1039>3.0.co;2-o]
56. Lo LY, Lin S, Hsu L. Motivation for online impulse buying: A two-factor theory perspective. Int J Inf Manag 2016 Oct;36(5):759-772. [doi: 10.1016/j.ijinfomgt.2016.04.012]
57. Lin S, Lo L. Motivation for using the social commerce website in the sharing economy: a two-factor theory perspective. 2016 Jul 1 Presented at: Pacific Asia Conference on Information Systems (PACIS 2016); June 27-July 1, 2016; Chiayi, Taiwan p. 1-11 URL: https://aisel.aisnet.org/pacis2016/108
58. Ou CX, Sia CL. To trust or to distrust, that is the question. Commun ACM 2009 May;52(5):135-139. [doi: 10.1145/1506409.1506442]
59. Chen Q, Clifford SJ, Wells WD. Attitude toward the site II: new information. J Advert Res 2002 Mar 01;42(2):33-45. [doi: 10.2501/JAR-42-2-33-45]

XSL•FO
RenderX

60. Ducoffe RH. How consumers assess the value of advertising. J Curr Issues Res Advert 1995 Mar;17(1):1-18. [doi: 10.1080/10641734.1995.10505022]
61. Goel L, Prokopec S. If you build it will they come?—An empirical investigation of consumer perceptions and strategy in virtual worlds. Electron Commer Res 2009 Mar 10;9(1-2):115-134. [doi: 10.1007/s10660-009-9031-7]
62. Hausman AV, Siekpe JS. The effect of web interface features on consumer online purchase intentions. J Bus Res 2009 Jan;62(1):5-13. [doi: 10.1016/j.jbusres.2008.01.018]
63. Kwak DHA, Ramamurthy KR, Nazareth D, Lee S. The moderating role of helper's high in anchoring process: An empirical investigation in the context of charity website design. Comput Hum Behav 2018 Jul;84:230-244. [doi: 10.1016/j.chb.2018.02.024]
64. Danish Habib M, Qayyum A. Cognitive emotion theory and emotion-action tendency in online impulsive buying behavior. J Manag Sci 2018 Mar;5(1):86-99. [doi: 10.20547/jms.2014.1805105]
65. Cimbalo RS, Beck KL, Sendziak DS. Emotionally toned pictures and color selection for children and college students. J Genet Psychol 1978 Dec;133(2):303-304. [doi: 10.1080/00221325.1978.10533389]
66. Hawlitschek F, Jansen L, Lux E, Teubner T, Weinhardt C. Colors and trust: the influence of user interface design on trust and reciprocity. New Jersey: IEEE Press; 2016 Jan Presented at: 2016 49th Hawaii International Conference on System Sciences (HICSS); January 5-8, 2016; Koloa, HI p. 590-599. [doi: 10.1109/hicss.2016.80]
67. Williams LE, Bargh JA. Experiencing physical warmth promotes interpersonal warmth. Science 2008 Oct 24;322(5901):606-607 [FREE Full text] [doi: 10.1126/science.1162548] [Medline: 18948544]
68. Fenko A, Schifferstein HN, Hekkert P. Looking hot or feeling hot: What determines the product experience of warmth? Mater Design 2010 Mar;31(3):1325-1331. [doi: 10.1016/j.matdes.2009.09.008]
69. Rizomyliotis I, Konstantoulaki K, Kostopoulos I. Reassessing the effect of colour on attitude and behavioural intentions in promotional activities: the moderating role of mood and involvement. Australasian Market J 2021 Jan 27;26(3):204-215. [doi: 10.1016/j.ausmj.2018.04.002]
70. Bellizzi JA, Hite RE. Environmental color, consumer feelings, and purchase likelihood. Psychol Market 1992 Sep;9(5):347-363. [doi: 10.1002/mar.4220090502]
71. Baxter SM, Ilicic J, Kulczynski A. Roses are red, violets are blue, sophisticated brands have a Tiffany Hue: the effect of iconic brand color priming on brand personality judgments. J Brand Manag 2017 Dec 12;25(4):384-394. [doi: 10.1057/s41262-017-0086-9]
72. Ettis SA. Examining the relationships between online store atmospheric color, flow experience and consumer behavior. J Retail Consum Serv 2017 Jul;37:43-55. [doi: 10.1016/j.jretconser.2017.03.007]
73. Ijzerman H, Semin GR. The thermometer of social relations: mapping social proximity on temperature. Psychol Sci 2009 Oct;20(10):1214-1220. [doi: 10.1111/j.1467-9280.2009.02434.x] [Medline: 19732385]
74. Citron FMM, Goldberg AE. Social context modulates the effect of physical warmth on perceived interpersonal kindness: a study of embodied metaphors. Lang Cognition 2014 Jan 16;6(1):1-11. [doi: 10.1017/langcog.2013.4]
75. Rai D, Lin CW, Yang CM. The effects of temperature cues on charitable donation. J Consum Market 2017 Jan 09;34(1):20-28. [doi: 10.1108/jcm-07-2015-1505]
76. Brasel SA, Gips J. Red Bull “Gives You Wings” for better or worse: A double-edged impact of brand exposure on consumer performance. J Consum Psychol 2011 Jan;21(1):57-64. [doi: 10.1016/j.jcps.2010.09.008]
77. Mudambi SM, Schuff D. What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quart 2010;34(1):185-200. [doi: 10.2307/20721420]
78. Kim JG, Kong HK, Karahalios K, Fu WT, Hong H. The power of collective endorsements: credibility factors in medical crowdfunding campaigns. 2016 May Presented at: 2016 CHI Conference on Human Factors in Computing Systems; May 2016; San Jose, CA p. 4538-4549. [doi: 10.1145/2858036.2858289]
79. Xu Q. Should I trust him? The effects of reviewer profile characteristics on eWOM credibility. Comput Hum Behav 2014 Apr;33:136-144. [doi: 10.1016/j.chb.2014.01.027]
80. Walther JB, Slovacek CL, Tidwell LC. Is a picture worth a thousand words? Commun Res 2016 Jun 30;28(1):105-134. [doi: 10.1177/009365001028001004]
81. Thabet M, Zghal M. An exploratory approach to the influence of perceived social presence on consumer trust in a website. J Int E-Bus Stud 2013 Jan 24(2013):1-20. [doi: 10.5171/2013.895661]
82. Calefato F, Lanubile F, Novielli N. The role of social media in affective trust building in customer–supplier relationships. Electron Commer Res 2015 Jul 16;15(4):453-482. [doi: 10.1007/s10660-015-9194-3]
83. Choi J, Li YJ, Rangan P, Yin B, Singh SN. Opposites attract: Impact of background color on effectiveness of emotional charity appeals. Int J Res Market 2020 Sep;37(3):644-660. [doi: 10.1016/j.ijresmar.2020.02.001]
84. Kumar Ranganathan S, Madupu V, Sen S, R. Brooks J. Affective and cognitive antecedents of customer loyalty towards e-mail service providers. J Serv Market 2013 May 24;27(3):195-206. [doi: 10.1108/08876041311330690]
85. Johnson D, Grayson K. Cognitive and affective trust in service relationships. J Bus Res 2005 Apr;58(4):500-507. [doi: 10.1016/S0148-2963(03)00140-1]
86. Kim JU, Kim WJ, Park SC. Consumer perceptions on web advertisements and motivation factors to purchase in the online shopping. Comput Hum Behav 2010 Sep;26(5):1208-1222. [doi: 10.1016/j.chb.2010.03.032]

XSL•FO
RenderX

87. Motoki K, Saito T, Nouchi R, Kawashima R, Sugiura M. Light colors and comfortable warmth: Crossmodal correspondences between thermal sensations and color lightness influence consumer behavior. Food Qual Pref 2019 Mar;72:45-55. [doi: 10.1016/j.foodqual.2018.09.004]
88. Friedrich T, Schlauderer S, Overhage S. The impact of social commerce feature richness on website stickiness through cognitive and affective factors: An experimental study. Electron Commerce Res Appl 2019 Jul;36:100861. [doi: 10.1016/j.elerap.2019.100861]
89. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
90. Nunnally J. Psychometric theory. 2nd edition. Hillsdale: McGraw-Hill Publishing; 1978.
91. Hayes A. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. New York, NY: Guilford Press; 2013.
92. Zhao Q, Chen CD, Wang JL, Chen P. Determinants of backers' funding intention in crowdfunding: Social exchange theory and regulatory focus. Telemat Informat 2017 Feb;34(1):370-384. [doi: 10.1016/j.tele.2016.06.006]
93. Morrow JJ, Hansen M, Pearson A. The cognitive and affective antecedents of general trust within cooperative organizations. J Manag Iss 2004;16(1):48-64.
94. Xiao Q. Understanding the asymmetric perceptions of smartphone security from security feature perspective: A comparative study. Telemat Informat 2021 May;58:101535. [doi: 10.1016/j.tele.2020.101535]
95. Social networking industry research report in 2019 (in Chinese). Jiguang. 2019 Apr 11. URL: https://www.jiguang.cn/reports/381/ [accessed 2020-04-26]

Abbreviations

ANCOVA: analysis of covariance
AVE: average variance extracted

Edited by G Eysenbach; submitted 06.11.20; peer-reviewed by Z Deng, J Mou; comments to author 24.11.20; revised version received 10.01.21; accepted 11.04.21; published 04.05.21.

Please cite as:
Zhang X, Hu W, Xiao Q
Influences of Medical Crowdfunding Website Design Features on Trust and Intention to Donate: Controlled Laboratory Experiment
J Med Internet Res 2021;23(5):e25554
URL: https://www.jmir.org/2021/5/e25554
doi:10.2196/25554
PMID:33944787

©Xing Zhang, Wenli Hu, Quan Xiao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2021. 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

Willingness to Adopt Health Information Among Social Question-and-Answer Community Users in China: Cross-sectional Survey Study

PengFei Li[1,2], MA; Lin Xu[1,2], MM; Tingting Tang[1,3], MA; Xiaoqian Wu[1,2], MM; Cheng Huang[1,2], MA

[1]Medical Informatics College, Chongqing Medical University, Chongqing, China

[2]Medical Data Science Academy, Chongqing Medical University, Chongqing, China

[3]The Children's Hospital of Chongqing Medical University, Chongqing, China

Corresponding Author:
Cheng Huang, MA
Medical Informatics College
Chongqing Medical University
No 1 Yixueyuan Road
Yuzhong District
Chongqing
China
Phone: 86 023 6848 0060
Email: huangcheng@cqmu.edu.cn

Abstract

Background: COVID-19 has spread around the world and has increased the public's need for health information in the process. Meanwhile, in the context of lockdowns and other measures for preventing SARS-CoV-2 spread, the internet has surged as a web-based resource for health information. Under these conditions, social question-and-answer communities (SQACs) are playing an increasingly important role in improving public health literacy. There is great theoretical and practical significance in exploring the influencing factors of SQAC users' willingness to adopt health information.

Objective: The aim of this study was to establish an extended unified theory of acceptance and use of technology model that could analyze the influence factors of SQAC users' willingness to adopt health information. Particularly, we tried to test the moderating effects that different demographic characteristics had on the variables' influences.

Methods: This study was conducted by administering a web-based questionnaire survey and analyzing the responses from a final total of 598 valid questionnaires after invalid data were cleaned. By using structural equation modelling, the influencing factors of SQAC users' willingness to adopt health information were analyzed. The moderating effects of variables were verified via hierarchical regression.

Results: Performance expectation (β=.282; *P*<.001), social influence (β=.238; *P*=.02), and facilitating conditions (β=.279; *P*=.002) positively affected users' willingness to adopt health information, whereas effort expectancy (*P*=.79) and perceived risk (*P*=.41) had no significant effects. Gender had a significant moderating effect in the structural equation model (*P*<.001).

Conclusions: SQAC users' willingness to adopt health information was evidently affected by multiple factors, such as performance expectation, social influence, and facilitating conditions. The structural equation model proposed in this study has a good fitting degree and good explanatory power for users' willingness to adopt health information. Suggestions were provided for SQAC operators and health management agencies based on our research results.

(J Med Internet Res 2021;23(5):e27811) doi:10.2196/27811

KEYWORDS

health information adoption; social question-and-answer community; structural equation model; Zhihu

XSL•FO
RenderX

Introduction

After the outbreak of COVID-19, the Chinese government implemented community isolation measures to control the spread of the epidemic, and because of these measures, the internet became the public's primary tool for searching for health information. Web-based knowledge-sharing platforms such as Zhihu (Zhihu Inc) and other social question-and-answer communities (SQACs) have been playing an increasingly important role in disseminating health information to the public. By 2019, Zhihu had more than 220 million registered users and had produced more than 28 million questions and 130 million answers. In recent months, the topic of COVID-19 has attracted more than 10,000 followers who had more than 13,000 related questions. Zhihu also has more than 20 million followers in their *Health* section and more than 750,000 health-related posts. We believe that exploring the willingness to adopt health information (WAHI) among users of SQACs could characterize community users' health information needs and behaviors. We believe that the findings of this study can promote public health literacy by improving the quality and efficiency of SQAC users' health information searches, adoption, and use.

Health information has flourished as a topic in recent information behavior literature. Many scholars have conducted in-depth research on health information resources [1-3], requesters [3-5], dissemination environments [6,7], and technology [8,9]. Specifically, after the epidemic of COVID-19 began, the research area of health information behavior attracted more attention [10-12]. Health information adoption is the final step in the health information diffusion path; it directly determines the effects of any information dissemination, and because of its significance, many researchers have investigated health information adoption among SQAC users. For instance, Diviani et al [13] found that low health literacy and related skills were negatively related to the ability to evaluate web-based health information and trust such information. Lee et al [14] investigated the level of seeking health-related information on the internet and how health literacy, access to technology, and sociodemographic characteristics impact behaviors related to seeking health-related information. Additionally, Ridout and Campbell [15] reviewed relevant research on young people's acquisition of health information from social networking sites and found that social networking site–based interventions were highly usable, engaging, and supportive for young people. More and more research findings have determined that further research on the influence factors of health information adoption intentions among SQAC users is necessary and meaningful.

SQACs are public social media platforms in which normal users both search for and share experiences and knowledge on any given topics [16]. These web-based communities have the characteristics of professionalism, interactivity, and open editing [17], and they have high potential influence on health information dissemination. Multimedia Appendix 1 shows a screenshot of the Zhihu website, and each part of the website is annotated in detail. In the existing research on SQACs, the two primary areas of focus have been platform design [18,19] and user behavior [16,17,20]. Jin et al [20] explored patient behaviors related to seeking health care information on SQACs based on dual-process theory and the knowledge adoption model, and Shi et al [21] compared the health information needs of Chinese and American patients with diabetes. The latter authors found that communities from different countries had different attributes. One commonality in extant health information adoption research is that it is grounded in the text information that already exists on social media platforms. There is far less research on SQAC users' health information behaviors based on users' own subjective self-reports.

Venkatesh et al [22] proposed the unified theory of acceptance and use of technology (UTAUT) by summarizing the advantages and disadvantages of other theoretical models in the field of technology acceptance. The core factors of the UTAUT include performance expectation (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FCs); gender, age, experience, and the voluntariness of use were used as moderators in the model. Due to the native characteristics of the UTAUT model, it has a stronger affinity and adaptability for most acceptance behavior research. Many scholars have embraced the UTAUT as a new theoretical blueprint in technology acceptance behavior research and have incorporated it into a vigorous acceptance behavior research stream. Bawack and Kamdjoug [23] conducted an empirical study on the acceptance of hospital information systems in limited-income countries based on the UTAUT. In a similar study based on the adjusted UTAUT, Nunes et al [24] found that gender, age, and other individual characteristics were moderating factors of users' willingness to adopt health applications. Abdelhamid [25] extended the UTAUT model and found that higher autonomy positively affected patients' willingness to exchange health information. Zhou et al [26] studied the mobile health (mHealth) information retrieval behavior of South African college students based on the UTAUT and found that perceived usefulness was the main influencing factor of their willingness to accept health information exchange. Liu et al [27,28] established a taxonomy of clinical decision support interventions based on the UTAUT model and tried to explore how patient care can be improved through clinical decision support. The UTAUT model has been used widely in the research area of users' acceptance behaviors. However, few scholars have used the UTAUT model to study SQAC users' WAHI. The goal of this study was to establish an effective model for analyzing SQAC users' WAHI based on the UTAUT model and to identify the influencing factors of users' WAHI.

Hypotheses and Modeling

PE Factor

PE is defined as the degree to which an individual believes that using a system will help them attain gains in job performance [22]. In the environment of SQACs, PE refers to users' perceptions of the benefits of adopting health information. Researchers measure PE with the following five structural variables: perceived usefulness, extrinsic motivation, job fit, relative advantage, and outcome expectations [29]. Based on the environment of SQACs, we proposed four additional PE items [29-31] on our questionnaire scale, as follows: (1) "the health information on Zhihu can play an important role in understanding and solving health-related problems," (2) "making

XSL•FO
RenderX

full use of the health information on Zhihu can help me better solve some problems," (3) "the health information on Zhihu has helped me or others around me improve our health," and (4) "the health information on Zhihu has a high reference value for my health decision-making." In general, researchers have shown that PE has positive effects on acceptance [32,33]. On the basis of these findings, we proposed the following research hypothesis: PE positively affects SQAC users' WAHI (H1).

EE Factor

In the UTAUT model, EE is defined as the degree of ease associated with the use of a system [22]. In this study, it refers to a user's perception of the difficulty of adopting relevant health information in SQACs and applying such information to practice. The following three constructs capture the concept of EE: the perceived ease of use, complexity, and ease of use [22]. Cimperman et al [34] and Adenuga et al [35] found that EE was a positive variable in models, and we proposed the following two scale items based on the previously mentioned studies [22,36] for our questionnaire: (1) "the health information on Zhihu is easy to understand" and (2) "the health recommendations on Zhihu are usually easier to implement." Users are more likely to adopt health information that is easy to understand and implement, and on the basis of these findings, we proposed the following hypothesis: EE positively affects SQAC users' WAHI (H2).

SI Factor

Venkatesh et al [22] defined SI as the degree to which an individual perceives that it is important for others to believe that they should use a new system, and SI can be represented by a subjective norm, social factors, and an image. The differences in health levels between other people before and after adopting health information will affect individuals' willingness to accept of health information. On the basis of previous research findings [36,37] and the three dimensions of SI measurement proposed by Venkatesh et al [22], we proposed the following three scale items: (1) "there are other people available to look up or solve health-related problems on Zhihu," (2) "other people around me have pushed health information from Zhihu to me," and (3) "Zhihu's image makes me feel that it will help me understand or solve health-related problems." On the basis of the research finding that SI has a positive effect on acceptance [38-40], we proposed the following hypothesis: SI positively affects SQAC users' WAHI (H3).

FCs Factor

FCs refer to the perceived (organizational, societal, etc) convenience of or support for adopting something new, such as a new technology [22]. In this study, we investigated the FCs of perceived health information convenience and support among SQAC users. Health information support in an SQAC environment mainly comprises the following three aspects: (1) community users' assistance with understanding relevant health information, (2) the convenience and ease of obtaining relevant health information from users, and (3) individual users' own knowledge and experiences. On the basis of these features [36], we proposed three questionnaire items, as follows: (1) "I have the resources, hardware and knowledge reserve to effectively use the health information on Zhihu," (2) "I can get help from others when I encounter problems while browsing and consulting health information on Zhihu," and (3) "seeking health information on Zhihu is one of the common ways for me to understand and solve health-related problems." Garavand et al [41] found that FCs had a positive effect on major students' adoption of mHealth apps. On the basis of these research findings, we proposed the following hypothesis: FCs positively affect SQAC users' WAHI (H4).

Perceived Risk

Perceived risk (PR) is one of the most important and widely used concepts in psychology, economics, and other fields. At present, the widely used measurement dimensions of PR include economic risk, time risk, information security risk, and health risk. PR is also one of the important determinants of health information adoption; the higher the perceived risk, the lower the willingness to adopt such information. A study [42] of Chinese patients' intention to adopt mHealth services showed that PR negatively affects respondents' trust and willingness to accept mHealth services. Although the UTAUT model integrates many variables from different classical models, it does not include the impact of PR on users' willingness to accept technology. Therefore, we decided to extend the PR variable in the model. Physiological risk, psychological risk, and time risk are the main risks for SQAC users' WAHI available on Zhihu. On the basis of the above findings, we proposed the following three items for our questionnaire scale: (1) "taking advice from relevant health information on Zhihu could cause physiological harm," (2) "taking advice from the health information on Zhihu could cause some psychological pressure," and (3) "the health information on Zhihu could be a useless waste of my time." Furthermore, we proposed the following hypothesis: PR negatively affects SQAC users' WAHI (H5).

Moderating Variables

Moderating variables such as gender, age, and the voluntariness of use play a very important role in the original UTAUT model [22]. The original UTAUT model is a widely used measurement model. However, when measuring and explaining a phenomenon with the aid of the UTAUT model, the variables in the model must be adjusted based on objective facts. When users decide whether to adopt certain health information, they are faced with high health costs, which are directly related to their own or other people's health. Compared to the intention to adopt information technology, the influence of the voluntariness of use variable on SQAC users' WAHI is less important. Therefore, the voluntariness of use variable was removed from the model. Demographic characteristics may have different levels of significance in terms of their effects on the WAHI. Haluza et al [43] found that sociodemographic attributes, including gender influence, not only affect private the web-based habits of users but might also affect the acceptance of health technologies and their professional use in a clinical setting. Tarver et al [44] found that age is one of the significant social-economic influencing factors of web-based health information communication. Irizarry et al [45] proposed that age and education level strongly influence patient users' interests and their ability to use patient portals. Based on existing relevant theories, we selected gender,

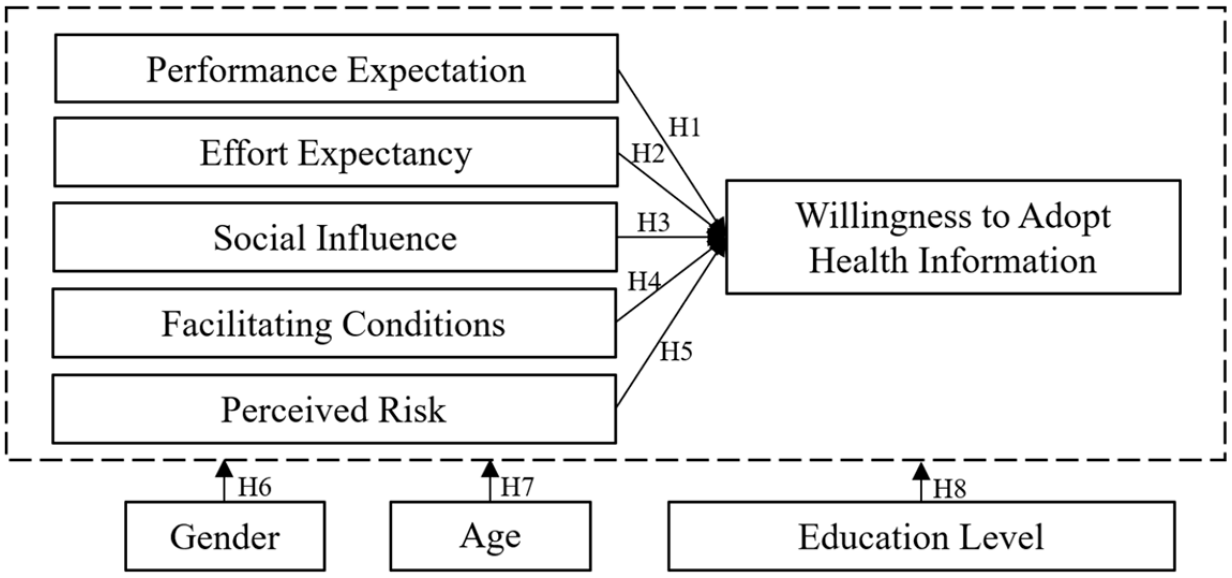
XSL•FO
RenderX

age, and education level as the three moderating variables in the model and proposed the following hypotheses: gender has a significant moderating effect on the influence of independent variables for adoption intention (H6), age has a significant moderating effect on the influence of independent variables for adoption intention (H7), and educational level has a significant moderating effect on the influence of independent variables for adoption intention (H8).

Model

Based on the structural characteristics of SQACs and SQAC users' adoption of health information, we reset the moderating variables of the UTAUT model and incorporated PR variables into the UTAUT model to construct the final model of SQAC users' WAHI. We present the model in Figure 1.

Figure 1. The model of social question-and-answer community users' willingness to adopt health information.



Methods

Participant Selection and Data Collection

We selected users who exhibited health information behaviors (health information browsing, commenting, searching, and other relevant behaviors) on Zhihu as the respondent sample for this study. On the basis of our literature review findings, we designed a web-based questionnaire survey scale. The scale was comprised of the items we previously outlined, and it was used to measure 19 indicators of 6 variables related to the WAHI among users of SQACs. All scale items were rated on 5-point Likert scales. We selected 22 postgraduates as presurvey participants to test the availability and quality of the questionnaire. Based on the presurvey results of the questionnaire, we deleted one of the index items for the variable EE (ie, "it often takes more time to retrieve health information in Zhihu"). After adjusting the questionnaire content and structure, the final scale was chosen via expert discussion. We included screening items in the final demographic indicators section of the questionnaire to ensure that all of the data we used came specifically from the Zhihu user group. We administered the web-based survey over 14 days between June 5 and 19, 2020, on the web-based questionnaire platform Wenjuanxing (Liepin Holdings Limited). The questionnaire was distributed by Zhihu users' forwarding the questionnaire link via WeChat (Tencent Holdings Limited). Before the survey process could advance, we presented the content of this study and required respondents to confirm their informed consent for participating further in this study. In total, data from 921 participants' were collected in this study. After filtering out the data of the nonusers of Zhihu and data with missing values and obvious errors, valid data from 598 participants were obtained, with an effective rate of 64.93%. A detailed list of questions can be found in Multimedia Appendix 2.

Statistical Analysis

We used Microsoft Excel for data cleaning and preprocessing before the data analysis. The statistical analysis tools we used in this study mainly included SPSS version 24.0 (IBM Corporation), Analysis of Moment Structures (AMOS) version 24.0 (IBM Corporation), and Process macro version 3.3 for SPSS [46]. Considering that the age distribution of Zhihu users was relatively concentrated, to avoid having a sample size of 0 for certain age groups and to increase the rate of unnecessary systematic errors, we grouped samples by age (20-year increments). With the structural equation model, we calculated the path coefficients between variables to verify the hypotheses. With Process, we conducted hierarchical regression to verify whether the moderating effects of each moderating variable in the model were significant. We set a *P* value of less than .05 as statistically significant.

Quality Control

Composite reliability and the Cronbach α are the most commonly used indicators of questionnaire reliability. As shown in Table 1, the Cronbach α and composite reliability values for each variable, except those for EE, were greater than .7, and the overall Cronbach α value for the questionnaire was .917. These results indicated the good reliability of the questionnaire that we developed for this study [47]. The validity testing of questionnaires entails testing two components—content and

XSL•FO
RenderX

structure validity. As our model and the index expression were verified by others many times, the questionnaire had high content validity. Structural validity consists of convergent and discriminant validities. Convergent validity requires the factor loading and average variance extraction (AVE) values for each index item to exceed .5. As shown in Table 1, all of the variables in this study exceeded this threshold, except for EE. Discriminant validity requires the correlation coefficient of each variable to be less than the square root of the AVE value of the variable itself, and Table 2 reflects that the questionnaire passed the discriminant validity test.

Table 1. The factor load, Cronbach α, average variance extraction (AVE), and composite reliability (CR) values of each variable.

Variables and indices	Factor load	Cronbach α	AVE	CR
PE[a]		.881	.636	.883
PE1	.804			
PE2	.819			
PE3	.824			
PE4	.787			
EE[b]		.450	.450	.621
EE1	.657			
EE3	.685			
SI[c]		.774	.531	.773
SI1	.745			
SI2	.688			
SI3	.752			
FC[d]		.760	.520	.764
FC1	.783			
FC2	.662			
FC3	.714			
PR[e]		.819	.603	.819
PR1	.789			
PR2	.812			
PR3	.725			
WAHI[f]		.852	.596	.854
WAHI1	.716			
WAHI2	.821			
WAHI3	.704			
WAHI4	.838			

[a]PE: performance expectation.

[b]EE: effort expectancy

[c]SI: social influence.

[d]FC: facilitating condition.

[e]PR: perceived risk.

[f]WAHI: willingness to adopt health information.

XSL•FO
RenderX

Table 2. Discriminant validity matrix.[a]

Variables	Performance expectation	Effort expectancy	Social influence	Facilitating condition	Perceived risk	WAHI[b]
Performance expectation	.797	N/A[c]	N/A	N/A	N/A	N/A
Effort expectancy	.535	.671	N/A	N/A	N/A	N/A
Social influence	.589	.507	.729	N/A	N/A	N/A
Facilitating condition	.609	.449	.590	.721	N/A	N/A
Perceived risk	.038	.289	.160	.144	.777	N/A
WAHI	.607	.444	.553	.576	.125	.772
AVE[d]	.636	.450	.531	.520	.603	.596

[a]The diagonal of the matrix is the square root of AVE of the corresponding variable.

[b]WAHI: willingness to adopt health information.

[c]N/A: not applicable.

[d]AVE: average variance extraction.

Results

Demographic Characteristics

The demographic characteristics of the participants in this study are shown in Table 3. Of the 598 respondents, 69.90% (n=419) were female and 94.15% (n=563) were aged 19 to 38 years. Just over three-quarters of the survey respondents (477/598, 79.77%) had an undergraduate or higher education.

Table 3. Statistical description of the sample.

Variables and categories	Value, n (%)
Gender	
Male	179 (29.93)
Female	419 (69.90)
Age (years)	
≤18	21 (3.51)
19-38	563 (94.15)
39-58	14 (2.34)
Education	
Senior high school and below	7 (1.17)
Junior college	10 (1.67)
Undergraduate	477 (79.77)
Master and above	104 (17.39)

Model Test

We completed the path verification of the model with AMOS 24.0 and SPSS 24.0, and Table 4 presents the model fitting findings. The model fit indices that we calculated were chi-square to df ratios, the root mean square error of approximation (RMSEA), the normed fit index, the relative fit index, the incremental fit index, the Tucker-Lewis index, and the cumulative fit index. Conventionally, model fit is considered good when the chi-square to df ratio is <3, the RMSEA is <0.08, and the normed fit index , relative fit index , incremental fit index, Tucker-Lewis index, and cumulative fit index are >0.9 [48,49].

XSL•FO
RenderX

Table 4. Model fitting.

Indices and values	Standard value	Fitting
Chi-square to df ratio (χ^2/df)		
2.954	<5	Acceptable
2.954	<3	Ideal
RMSEA[a]		
0.057	<0.08	Acceptable
0.057	<0.05	Ideal
Normed fit index		
0.929	>0.9	Ideal
Relative fit index		
0.912	>0.9	Ideal
Incremental fit index		
0.952	>0.9	Ideal
Tucker-Lewis index		
0.940	>0.9	Ideal
Cumulative fit index		
0.952	>0.9	Ideal

[a]RMSEA: root mean square error of approximation.

The structural equation model and the path coefficients are shown in Table 5. Paths "EE to WAHI" (*P*=.79) and "PR to WAHI" (*P*=.41) were not significant, whereas paths "PE to WAHI" (β=.282; *P*<.001), "SI to WAHI" (β=.238; *P*=.02), and "FC to WAHI" (β=.279; *P*=.002) were significant. Our findings supported H1, H3, and H4 but failed to support H2 and H5.

Table 5. Path test of the structural equation.

Hypotheses	Path	Unstandardized estimates	Standardized estimates (SE[a])	Critical ratio	*P* value	Results
H1	PE[b] to WAHI[c]	.280	.282 (.084)	3.314	<.001	Accept
H2	EE[d] to WAHI	.036	.027 (.141)	0.256	.79	Reject
H3	SI[e] to WAHI	.224	.238 (.098)	2.296	.02	Accept
H4	FC[f] to WAHI	.262	.279 (.085)	3.080	.002	Accept
H5	PR[g] to WAHI	.030	.032 (.036)	0.825	.41	Reject

[a]SE: standard error.

[b]PE: performance expectation.

[c]WAHI: willingness to adopt health information.

[d]EE: effort expectancy.

[e]SI: social influence.

[f]FC: facilitating condition.

[g]PR: perceived risk.

Moderating Effect Test

With Process macro version 3.3 for SPSS, we completed the testing of the moderating effects of the moderating variables in the model, and we present the specific significance of moderating effects in Table 6. The results in Table 6 indicate that gender significantly moderated the effect that PE had on SQAC users' WAHI. This supported H6. However, H7 and H8 were not supported. Table 7 shows the model parameters from before and after we incorporated gender as a moderating variable. With regard to model 1, Table 7 presents the standardized coefficients of each variable and the corresponding R^2 and *F* test values for when the moderating variables were not included. With regard to model 2, Table 7 reflects the parameter changes that occurred after we introduced gender as a moderating variable. Compared to model 1, model 2's explanatory power for SQAC users' WAHI improved after we incorporated the different moderating variables. Figure 2 shows a visual representation of the moderating effects in this study.

XSL•FO
RenderX

The slope of Figure 2 indicates the effect that PE has on SQAC users' WAHI. A larger slope means that the model is more sensitive to the WAHI. We found that the slope of the male sample was considerably larger than that of the female sample (Figure 2). Therefore, we believe that the male group had more obvious fluctuations than the female group.

Table 6. The significance of moderating effects.

Path	Gender	Age	Education level
PE[a] to WAHI[b]	√[c]		
SI[d] to WAHI			
FC[e] to WAHI			

[a]PE: performance expectation.

[b]WAHI: willingness to adopt health information.

[c]The moderating effect was significant at the .05 level.

[d]SI: social influence.

[e]FC: facilitating condition.

Table 7. Hierarchical regression test of moderating effects.

Index	Model 1[a]			Model 2[b]		
	B	*t* test (*df*)	*P* value	B	*t* test (*df*)	*P* value
Performance expectation	0.331	8.126 (597)	<.001	0.375	8.123 (597)	<.001
Social influence	0.210	5.250 (597)	<.001	0.190	5.029 (597)	<.001
Facilitating conditions	0.251	6.171 (597)	<.001	0.249	6.179 (597)	<.001
Gender	N/A[c]	N/A	N/A	0.094	1.913 (597)	<.001
Interaction item[d]	N/A	N/A	N/A	−0.170	−2.390 (597)	.02

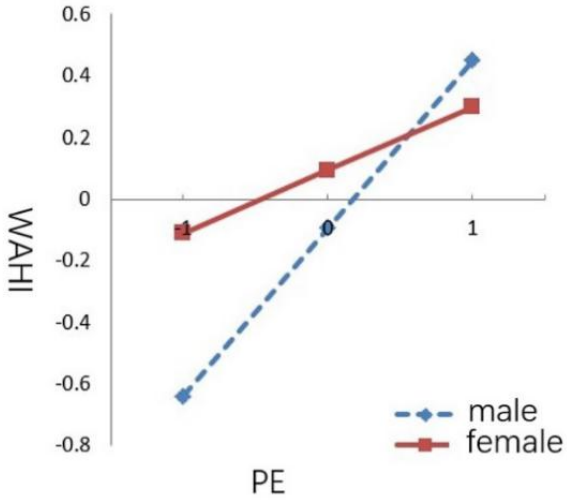
[a]Model 1 had an R^2 value of 0.461 (P<.001) and an F test value ($F_{3,594}$) of 169.643.

[b]Model 2 had an R^2 value of 0.470 (P<.001) and an F test value ($F_{5,592}$) of 104.804.

[c]N/A: not applicable.

[d]The interaction item for the gender and performance expectation.

Figure 2. The moderating effect of gender on the "PE to WAHI" path. PE: performance expectation; WAHI: willingness to adopt health information.



Discussion

Principal Findings

Since the outbreak of COVID-19, the role of SQACs in the public's access to health information has grown in prominence. This study was an analysis of the influencing factors of SQAC users' WAHI during the COVID-19 pandemic. We conducted measurements by using a questionnaire that comprised items grounded in the UTAUT and its individual components. Based on our results, PE (P<.001) and SI (P=.02) had significant positive effects on the WAHI. PE reflects users' expectations

XSL•FO
RenderX

for improved health if they adopt health information, and the WAHI increases in a community if the community users believe that the information is helpful to them. However, in contrast with PE, SI affected WAHI through the people around the users. Improved health among individuals who adopt health information will encourage SQAC users to adopt the health information that is recommended by these individuals. Abdelhamid [25] found that perceived usefulness and SI were the most important factors affecting health information exchange between patients, and our findings were consistent with that conclusion.

Under the premises of compliance and legality, SQAC operators can make full use of the traces of users in SQACs (eg, browsing content, page stay times, likes, collections, content, and the theme of private messages). With the help of this kind of information, SQAC operators could enhance the strength of relationships among different users who pay attention to the same health topics. With the epidemic of COVID-19, people's health concerns have become more focused. This provides a rare opportunity for SQAC operators to improve users' relationship strength. Furthermore, actively guiding users to share their understanding of and experience with health information is another effective method for improving users' PEs and SI. By using these methods, we can further enhance users' WAHI and users' health levels.

FCs positively affected SQAC users' WAHI. FCs are an integration of users' peripheral auxiliary functions in the process of adopting health information. They are formed based on individuals' levels of comprehension, the convenience of retrieving information from platforms, and other users' help in understanding health information. When the effectiveness of this auxiliary role improves, the process of SQAC users' adoption of health information becomes smoother, and users' confidence in adopting health information increases. Through this virtuous cyclical process, health information is continuously shared, exchanged, and adopted, and improving the health information retrieval mechanism can increase the quality of the health information retrieved.

EE (P=.79) and PR (P=.41) had no significant effects on SQAC users' WAHI. We believe that it was the interaction of many factors that led to this result. The native environment and characteristics of the original UTAUT could explain why EE had no significant effects on WAHI. Compared to the cost of users' adoption of information technology or information systems in the study conducted by Venkatesh et al [22], during the establishment of the original UTAUT model, the cost of SQAC users' adoption of health information was lower. To some extent, if the cost of health information adoption is low, the significance level of EE will also be low. However, risk comes from the unknown. Sudhakar et al [50] found that lower education levels were significantly related to lower health literacy. For most SQAC users, higher education means higher information literacy. In this study, irrespective of the total number of SQAC users or the sample size, users' demographic characteristics reflected a higher educational level, which is important for understanding and using health information. Coughlin et al [51] found that persons with limited health literacy are less likely to use patient web portals. We believe that SQAC users who adopt health information and have lower educational levels face more difficulties and risks than those with higher educational levels. However, there could also be deeper reasons for why EE and PR did not significantly affect SQAC users' WAHI. Our inconclusive results require us to conduct more in-depth and detailed research.

As shown in Tables 6 and 7, the moderating effect of gender on path "PE to WAHI" (P=.02) was significant, and Figure 2 shows that the WAHI among men was more sensitive to PE than the WAHI among women. Baumann et al [52] found that gender had different moderating effects on the influencing factors of web users' health information search behaviors. Through empirical research, Ek [53] found that men received far less informal health information from family members, friends, and colleagues than women. Therefore, the PEs of men are more effective in increasing their WAHI than the PEs of women. SQAC operators should pay more attention to the moderating role of gender in the "PE to WAHI" path when attempting to identify optimal health information dissemination schemes. Different coping strategies should be implemented for users of different genders. However, we can effectively improve the WAHI among SQAC users by differentiating the information that they receive based on user gender.

Contributions and Limitations

In this study, we proposed a model of the WAHI among users of the Zhihu SQAC that was based on the modification of the native UTAUT, and we found good explanatory power for our model. We also analyzed the different influencing factors of the WAHI among Zhihu users. PE, SI, and FCs were the primary influencing factors, and the effect of PE differed according to gender. We proposed several suggestions and measures that can be implemented based on our research findings in this study.

Although we strove to be rigorous, there were still several limitations in this study. First, this was a questionnaire survey based on the subjective cognition of SQAC users; thus, we could not avoid the interference of various subjective factors associated with self-reporting. Second, this was a cross-sectional study, and as such, it was impossible to observe changes in SQAC users' WAHI over time. Third, although we ensured that the sample was as representative as possible, there were still some inevitable systematic errors. Fourth, although SQACs are gradually becoming an indispensable platform that users can use to obtain health information, most users are still using search engines and other methods to obtain such information, and we did not adequately explain the interactions among these different sources of health information and users' willingness to adopt such information. Fifth, the model established in this study is a limited extension of the UTAUT model, which cannot cover all of the influencing factors of the dependent variables. Other variables such as information quality, trust, and medical experience will be modeled and studied as the focus in follow-up research. In addition, there were still several limitations in our choices for variable indices. After considering the efficiency of this study, we excluded some indicators that we subjectively considered unimportant, but whether the inclusion of these indicators would enhance the explanatory power of the model remains to be further studied. Finally, we only attempted to

XSL•FO
RenderX

identify the moderating effects of demographic characteristics. Therefore, only gender, age, and education level were selected for verification. Whether other demographic indicators have a significant impact needs to be further verified. In spite of the above limitations, the conclusions and suggestions of this study can be used as references by relevant health management agencies.

Conclusions

We constructed a UTAUT-based model to explain the WAHI among users of the Zhihu SQAC during the COVID-19 pandemic. We tested our hypotheses by using data from a survey (which we administered on the internet) that we analyzed via structural equation modelling. The results showed that PE, SI, and FCs had positive effects on SQAC users' WAHI; EE and PR did not affect users' WAHI. We also found that gender (P=.02) had a significant moderating effect in the model. We hold the opinion that enhancing the strength of user relationships and improving users' experiences with SQAC platforms are the most useful methods for improving the WAHI among users of SQACs. In addition, we need to encourage all users to improve their health information literacy. Although there are limitations in this study, SQAC operators, researchers, and policy makers can refer to our results to guide policy decisions.

Acknowledgments

This study was supported by the Western China Social Sciences Program "Research on Health Information Transmission Path and Diffusion Model Based on Mobile Internet" (item number: 16XTQ012). We appreciate the help of all of the participants who participated in the questionnaire survey. In addition, we are also very grateful to all reviewers for their careful examination of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Notes on Zhihu (Zhihu Inc).
[PNG File , 463 KB - jmir_v23i5e27811_app1.png]

Multimedia Appendix 2
Questionnaire on social question-and-answer community users' willingness to share health information.
[DOCX File , 27 KB - jmir_v23i5e27811_app2.docx]

References

1. Badu E, Mensah I, Gyamfi N, Agyei-Okyere E, Eric A, Adusei-Nkrumah J. Knowledge and sources of accessing sexual and reproductive health information among visually impaired women in Ghana. BMC Res Notes 2019 Aug 20;12(1):529 [FREE Full text] [doi: 10.1186/s13104-019-4568-6] [Medline: 31429803]
2. Ibegbulam IJ, Akpom CC, Enem FN, Onyam DI. Use of the internet as a source for reproductive health information seeking among adolescent girls in secondary schools in Enugu, Nigeria. Health Info Libr J 2018 Dec;35(4):298-308. [doi: 10.1111/hir.12242] [Medline: 30426642]
3. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]
4. Gonsalves L, Hindin MJ, Bayer A, Carcamo CP, Gichangi P, Habib N, et al. Protocol of an open, three-arm, individually randomized trial assessing the effect of delivering sexual and reproductive health information to young people (aged 13-24) in Kenya and Peru via mobile phones: adolescent/youth reproductive mobile access and delivery initiative for love and life outcomes (ARMADILLO) study stage 2. Reprod Health 2018 Jul 11;15(1):126 [FREE Full text] [doi: 10.1186/s12978-018-0568-6] [Medline: 29996854]
5. Patterson SP, Hilton S, Flowers P, McDaid LM. What are the barriers and challenges faced by adolescents when searching for sexual health information on the internet? Implications for policy and practice from a qualitative study. Sex Transm Infect 2019 Sep;95(6):462-467 [FREE Full text] [doi: 10.1136/sextrans-2018-053710] [Medline: 31040251]
6. Chen D, Zhang R, Feng J, Liu K. Fulfilling information needs of patients in online health communities. Health Info Libr J 2020 Mar;37(1):48-59. [doi: 10.1111/hir.12253] [Medline: 31090185]
7. Li Y, Song Y, Zhao W, Guo X, Ju X, Vogel D. Exploring the role of online health community information in patients' decisions to switch from online to offline medical services. Int J Med Inform 2019 Oct;130:103951. [doi: 10.1016/j.ijmedinf.2019.08.011] [Medline: 31473534]
8. Celis J, Baum A, Bordoni MVG, Alassia L, Stieben A, Franco M, et al. Workforce development strategy for health information system implementation at the public health system of Buenos Aires. Stud Health Technol Inform 2019 Aug 21;264:1905-1906. [doi: 10.3233/SHTI190706] [Medline: 31438400]

XSL•FO
RenderX

9. Vest JR, Ben-Assuli O. Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. Int J Med Inform 2019 Sep;129:205-210. [doi: 10.1016/j.ijmedinf.2019.06.013] [Medline: 31445257]
10. Di Renzo L, Gualtieri P, Pivari F, Soldati L, Attinà A, Cinelli G, et al. Eating habits and lifestyle changes during COVID-19 lockdown: an Italian survey. J Transl Med 2020 Jun 08;18(1):229 [FREE Full text] [doi: 10.1186/s12967-020-02399-5] [Medline: 32513197]
11. Li G, Tang D, Song B, Wang C, Qunshan S, Xu C, et al. Impact of the COVID-19 pandemic on partner relationships and sexual and reproductive health: Cross-sectional, online survey study. J Med Internet Res 2020 Aug 06;22(8):e20961 [FREE Full text] [doi: 10.2196/20961] [Medline: 32716895]
12. Rovetta A, Bhagavathula AS. COVID-19-related web search behaviors and infodemic attitudes in Italy: Infodemiological study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
13. Diviani N, van den Putte B, Giani S, van Weert JC. 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]
14. Lee HY, Jin SW, Henning-Smith C, Lee J, Lee J. Role of health literacy in health-related information-seeking behavior online: Cross-sectional Study. J Med Internet Res 2021 Jan 27;23(1):e14088 [FREE Full text] [doi: 10.2196/14088] [Medline: 33502332]
15. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: Systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
16. Charlie AM, Gao Y, Heller SL. What do patients want to know? Questions and concerns regarding mammography expressed through social media. J Am Coll Radiol 2018 Oct;15(10):1478-1486. [doi: 10.1016/j.jacr.2017.09.020] [Medline: 29221997]
17. Alasmari A, Zhou L. Int J Med Inform 2019 Nov;131:103958. [doi: 10.1016/j.ijmedinf.2019.103958] [Medline: 31521012]
18. Khansa L, Ma X, Liginlal D, Kim SS. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. J Manag Inf Syst 2015 Aug 28;32(2):162-203. [doi: 10.1080/07421222.2015.1063293]
19. Liu Y, Lin Z, Zheng X, Chen D. Incorporating social information to perform diverse replier recommendation in question and answer communities. Journal of Information Science 2015 Jul 21;42(4):449-464. [doi: 10.1177/0165551515592093]
20. Jin J, Yan X, Li Y, Li Y. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
21. Shi J, Qian Y, Li C, Zhou L, Zhang B. Comparing information needs of diabetes patients in Chinese and American health communities of questions and answers. Stud Health Technol Inform 2019 Aug 21;264:1031-1035. [doi: 10.3233/SHTI190381] [Medline: 31438081]
22. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
23. Bawack RE, Kamdjoug JRK. Adequacy of UTAUT in clinician adoption of health information systems in developing countries: The case of Cameroon. Int J Med Inform 2018 Jan;109:15-22. [doi: 10.1016/j.ijmedinf.2017.10.016] [Medline: 29195701]
24. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: Examining key determinants and moderators. Front Psychol 2019 Dec 10;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
25. Abdelhamid M. Greater patient health information control to improve the sustainability of health information exchanges. J Biomed Inform 2018 Jul;83:150-158 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.002] [Medline: 29894811]
26. Zhou LL, Owusu-Marfo J, Antwi HA, Antwi MO, Kachie ADT, Ampon-Wireko S. "Assessment of the social influence and facilitating conditions that support nurses' adoption of hospital electronic information management systems (HEIMS) in Ghana using the unified theory of acceptance and use of technology (UTAUT) model". BMC Med Inform Decis Mak 2019 Nov 21;19(1):230 [FREE Full text] [doi: 10.1186/s12911-019-0956-z] [Medline: 31752840]
27. Liu S, Reese TJ, Kawamoto K, Fiol GD, Weir C. Toward optimized clinical decision support: A theory-based approach. 2020 Presented at: 2020 IEEE International Conference on Healthcare Informatics (ICHI); November 30 to December 3, 2020; Oldenburg, Germany. [doi: 10.1109/ichi48887.2020.9374346]
28. Liu S, Reese TJ, Kawamoto K, Fiol GD, Weir C. A systematic review of theoretical constructs in CDS literature. BMC Med Inform Decis Mak 2021 Mar 17;21(1):102 [FREE Full text] [doi: 10.1186/s12911-021-01465-2] [Medline: 33731089]
29. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Manage Sci 2000 Feb 01;46(2):169-332. [doi: 10.1287/mnsc.46.2.186.11926]
30. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technologies: A dual-factor model. 2007 Presented at: 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07); January 3-6, 2007; Waikoloa, HI, USA. [doi: 10.1109/hicss.2007.437]
31. 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]
32. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]

XSL•FO
RenderX

33. Hennemann S, Witthöft M, Bethge M, Spanier K, Beutel ME, Zwerenz R. Acceptance and barriers to access of occupational e-mental health: cross-sectional findings from a health-risk population of employees. Int Arch Occup Environ Health 2018 Apr;91(3):305-316. [doi: 10.1007/s00420-017-1280-5] [Medline: 29189895]
34. Cimperman M, Brenčič MM, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
35. Adenuga KI, Iahad NA, Miskon S. Towards reinforcing telemedicine adoption amongst clinicians in Nigeria. Int J Med Inform 2017 Aug;104:84-96. [doi: 10.1016/j.ijmedinf.2017.05.008] [Medline: 28599820]
36. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep 01;2(3):173-239. [doi: 10.1287/isre.2.3.192]
37. Gefen D, Karahanna E, Straub DW. Trust and TAM in online shopping: An integrated model. MIS Q 2003 Mar;27(1):51-90. [doi: 10.2307/30036519]
38. Kim S, Lee KH, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
39. Lu X, Zhang R, Zhu X. An empirical study on patients' acceptance of physician-patient interaction in online health communities. Int J Environ Res Public Health 2019 Dec 12;16(24):5084 [FREE Full text] [doi: 10.3390/ijerph16245084] [Medline: 31842465]
40. Quaosar GMAA, Hoque MR, Bao Y. Investigating factors affecting elderly's intention to use m-Health services: An empirical study. Telemed J E Health 2018 Apr;24(4):309-314. [doi: 10.1089/tmj.2017.0111] [Medline: 28976824]
41. Garavand A, Samadbeik M, Nadri H, Rahimi B, Asadi H. Effective factors in adoption of mobile health applications between medical sciences students using the UTAUT model. Methods Inf Med 2019 Nov;58(4-05):131-139. [doi: 10.1055/s-0040-1701607] [Medline: 32170717]
42. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: Empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
43. Haluza D, Wernhart A. Does gender matter? Exploring perceptions regarding health technologies among employees and students at a medical university. Int J Med Inform 2019 Oct;130:103948. [doi: 10.1016/j.ijmedinf.2019.08.008] [Medline: 31442846]
44. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: Trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
45. 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]
46. Hayes AF, Rockwood NJ. Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. Behav Res Ther 2017 Nov;98:39-57. [doi: 10.1016/j.brat.2016.11.001] [Medline: 27865431]
47. Segars AH. Assessing the unidimensionality of measurement: a paradigm and illustration within the context of information systems research. Omega (Westport) 1997 Feb;25(1):107-121. [doi: 10.1016/s0305-0483(96)00051-5]
48. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
49. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
50. Sudhakar S, Aebi ME, Burant CJ, Wilson B, Wenk J, Briggs FBS, et al. Health literacy and education level correlates of participation and outcome in a remotely delivered epilepsy self-management program. Epilepsy Behav 2020 Jun;107:107026. [doi: 10.1016/j.yebeh.2020.107026] [Medline: 32249034]
51. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
52. 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]
53. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745. [doi: 10.1093/heapro/dat063] [Medline: 23985248]

Abbreviations

AMOS: Analysis of Moment Structures
AVE: average variance extraction
EE: effort expectancy

XSL•FO
RenderX

FC: facilitating condition
mHealth: mobile health
PE: performance expectation
PR: perceived risk
RMSEA: root mean square error of approximation
SI: social influence
SQAC: social question-and-answer community
UTAUT: unified theory of acceptance and use of technology
WAHI: willingness to adopt health information

Edited by C Basch; submitted 08.02.21; peer-reviewed by N Iahad, S Liu; comments to author 01.03.21; revised version received 10.04.21; accepted 08.05.21; published 21.05.21.

Please cite as:
Li P, Xu L, Tang T, Wu X, Huang C
Willingness to Adopt Health Information Among Social Question-and-Answer Community Users in China: Cross-sectional Survey Study
J Med Internet Res 2021;23(5):e27811
URL: https://www.jmir.org/2021/5/e27811
doi:10.2196/27811
PMID:33970865

©PengFei Li, Lin Xu, Tingting Tang, Xiaoqian Wu, Cheng Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.05.2021. 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 Clinical Prediction Model to Predict Heparin Treatment Outcomes and Provide Dosage Recommendations: Development and Validation Study

Dongkai Li[1*], MD, PhD; Jianwei Gao[2*], PhD; Na Hong[2*], PhD; Hao Wang[1*], MD, PhD; Longxiang Su[1], MD, PhD; Chun Liu[2], PhD; Jie He[2], PhD; Huizhen Jiang[3], MSc; Qiang Wang[2], MSc; Yun Long[1], MD, PhD; Weiguo Zhu[3,4], MD, PhD

[1]Department of Critical Care Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

[2]Digital Health China Technologies Co. Ltd., Beijing, China

[3]Department of Information Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[4]Department of Primary Care and Family Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Weiguo Zhu, MD, PhD
Department of Information Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College
No.1 Shuaifuyuan, Dongcheng District
Beijing,
China
Phone: 86 69154149
Email: zhuwg@pumch.cn

Abstract

Background: Unfractionated heparin is widely used in the intensive care unit as an anticoagulant. However, weight-based heparin dosing has been shown to be suboptimal and may place patients at unnecessary risk during their intensive care unit stay.

Objective: In this study, we intended to develop and validate a machine learning–based model to predict heparin treatment outcomes and to provide dosage recommendations to clinicians.

Methods: A shallow neural network model was adopted in a retrospective cohort of patients from the Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III) database and patients admitted to the Peking Union Medical College Hospital (PUMCH). We modeled the subtherapeutic, normal, and supratherapeutic activated partial thromboplastin time (aPTT) as the outcomes of heparin treatment and used a group of clinical features for modeling. Our model classifies patients into 3 different therapeutic states. We tested the prediction ability of our model and evaluated its performance by using accuracy, the kappa coefficient, precision, recall, and the F1 score. Furthermore, a dosage recommendation module was designed and evaluated for clinical decision support.

Results: A total of 3607 patients selected from MIMIC III and 1549 patients admitted to the PUMCH who met our criteria were included in this study. The shallow neural network model showed results of F1 scores 0.887 (MIMIC III) and 0.925 (PUMCH). When compared with the actual dosage prescribed, our model recommended increasing the dosage for 72.2% (MIMIC III, 1240/1718) and 64.7% (PUMCH, 281/434) of the subtherapeutic patients and decreasing the dosage for 80.9% (MIMIC III, 504/623) and 76.7% (PUMCH, 277/361) of the supratherapeutic patients, suggesting that the recommendations can contribute to clinical improvements and that they may effectively reduce the time to optimal dosage in the clinical setting.

Conclusions: The evaluation of our model for predicting heparin treatment outcomes demonstrated that the developed model is potentially applicable for reducing the misdosage of heparin and for providing appropriate decision recommendations to clinicians.

(J Med Internet Res 2021;23(5):e27118) doi:10.2196/27118

XSL•FO
RenderX

KEYWORDS

outcome prediction; clinical decision support; dosage recommendation; machine learning; intensive care unit

Introduction

Existing rule-based protocols guide clinicians to initiate, modulate, or terminate a certain treatment procedure according to evidence-based clinical guidelines or best practices [1]. However, in real clinical scenarios, dynamic changes occur continuously in individual patients with complex diseases and physiological situations, which frequently exceed the scope of the typical model described in the guidelines [2]. In particular, in critical care settings, there is a decision-making dilemma between evidence-based medicine and individualized medicine due to the lack of high-quality evidence and the urgency for the accuracy and effectiveness of the treatment [3]. Fortunately, in the era of big data and artificial intelligence and based on progress in data acquisition, integration, and application in critical care medicine, machine learning techniques sometimes can help with diagnosis, treatment, and prediction in intensive care units (ICUs) [4-6].

Heparin (or unfractionated heparin, UFH) infusion is one example of a medication for which retrospective data have been proven to provide valuable support in clinical settings, especially in critical settings in which patients are vulnerable to thromboembolism or hemorrhage or to severe complications caused by these conditions [7]. For decades, UFH dosing has been based solely on a patient's weight: a weight-based heparin dosing nomogram is the standard practice for the application of UFH. The heparin nomogram mainly consists of an empirical initial loading dose followed by a step-by-step modulation according to a series of blood clotting parameters that are monitored every 4-6 hours [8,9]. The actual optimal UFH dosage varies widely among patients with different physiological situations, and for these patients, the time to optimal dosage may be prolonged. Besides, adverse events associated with supratherapeutic or subtherapeutic anticoagulation, such as hemorrhagic tendency or thrombophilia, may occur due to the intravenous heparin's narrow therapeutic window [7]. Therefore, retrospective analysis starts by extracting sequential dose response data as well as the concurrent laboratory and other clinical data, which may greatly contribute to the improvement on the feedback delay of UFH dosage optimization [10]. In our recent study [11], based on the public Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC III) and electronic ICU databases [12,13], we compared several common models for predicting the effects of heparin treatment and showed that machine learning–based models, which outperformed the standard practices, can be used to optimize and personalize heparin dosing to improve patient safety.

Although our UFH model has been evaluated and validated using public databases [10,14], it has not been validated in our local clinical setting, especially for step-by-step modulation. In this study, we extend the validation and application of the previously optimized UFH model to a local clinical database in the Department of Critical Care Medicine of Peking Union Medical College Hospital (PUMCH), which is a complementary ICU in a tertiary hospital in China. Furthermore, UFH dosage recommendation based on the machine learning model was also performed for both the MIMIC and PUMCH databases, with the goal of effectively reducing the time to optimal dosage in the clinical setting. Finally, to further promote understanding of the data model and algorithm, we performed a featured importance analysis in both databases.

Methods

Data Sets

To evaluate the adaptability of the predictive model and to implement external validation, we employed patient data, including heparin treatments from 2 databases: the MIMIC III database and the PUMCH ICU database. The MIMIC III database is a free and open intensive care medical data set published by the Computational Physiology Laboratory of the Massachusetts Institute of Technology, the Beth Israel Deaconess Medical Center, and Philips Healthcare. It contains real medical data from more than 50,000 adult patients in the ICU at the Beth Israel Deaconess Medical Center between 2001 and 2012 [8]. The PUMCH ICU database comprises the complete clinical data of patients admitted to the PUMCH ICU with a retrospective cohort of more than 20,000 ICU patients between 2013 and 2019.

Definition of Heparin Treatment Outcomes

We classified patients as subtherapeutic, normal therapeutic, and supratherapeutic according to their therapeutic activated partial thromboplastin time (aPTT) values after heparin treatment. We used the average aPTT value from 8 h to 24 h after the initial heparin infusion as the therapeutic aPTT value. Figure 1 shows the distributions of the aPTT values in the 2 data sets. Due to differences in the regions, patient characteristics, and treatment plans such as the step-by-step treatment pattern of PUMCH, the observed therapeutic aPTT distributions are quite different. Based on a previous study [14] and suggestions from clinicians, we adopted different definition ranges for the 2 data sets (Table 1). We then labelled each patient record with 1 of the 3 labels (subtherapeutic, normal therapeutic, and supratherapeutic), thereby converting the clinical outcome prediction task into a ternary classification task.

XSL•FO
RenderX

Figure 1. The therapeutic activated partial thromboplastin time distributions in the 2 data sets. aPTT: activated partial thromboplastin time; MIMIC III: Multiparameter Intelligent Monitoring in Intensive Care III; PUMCH: Peking Union Medical College Hospital.

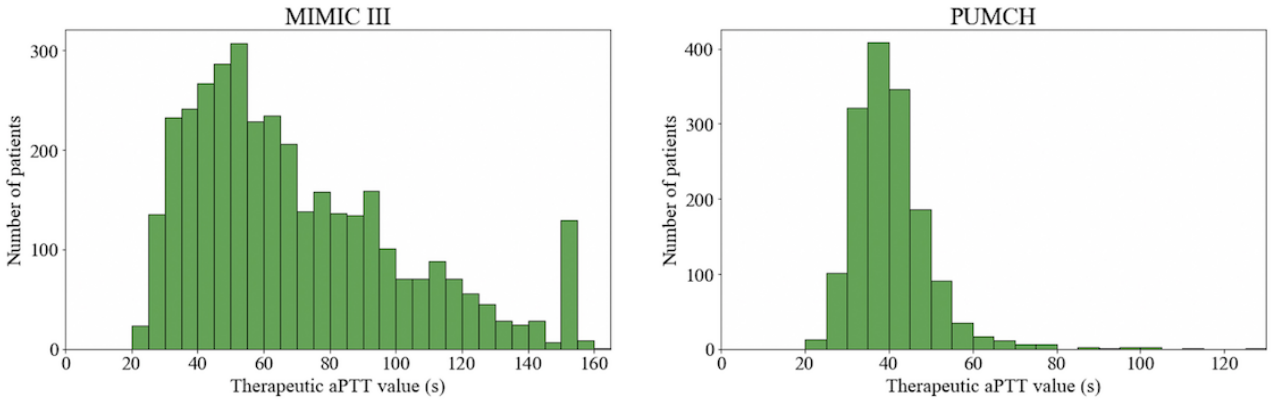


Table 1. Ranges of the therapeutic activated partial thromboplastin time classification of the 2 data sets.

Activated partial thromboplastin time (time in seconds)	Multiparameter Intelligent Monitoring in Intensive Care III data set	Peking Union Medical College Hospital data set
Subtherapeutic activated partial thromboplastin time (s)	0-60	0-35
Normal therapeutic activated partial thromboplastin time (s)	60-100	35-45
Supratherapeutic activated partial thromboplastin time (s)	>100	>45

Feature Selection

According to evidence from related studies and experience from clinical practice, various clinical features affect heparin treatment outcomes [14], that is, the therapeutic aPTT. For example, creatinine in the blood is almost entirely filtered into the urine via glomerular filtration and its concentration is stable under normal circumstances. Therefore, creatinine concentration in the blood can be used as an indicator of renal function because it reflects the filtration function of glomeruli. Measurements of renal, hepatic, cardiac, and coagulation functions were included as features of interest. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) concentrations in the blood are sensitive to hepatocellular damage, and their ratio is an important indicator of liver function. Sequential organ failure assessment (SOFA) scores were included. Furthermore, the total heparin dosage, defined as the sum of the heparin doses administered within 8 h of the initial heparin infusion, is also considered as the affected factor of heparin treatment outcomes. Therefore, to optimize the model predictions, we selected clinical features of interest from the 2 data sets, including sex, ethnicity, admission type, age, weight, initial aPTT, creatinine, AST/ALT ratio (we used the ALT value instead for PUMCH since AST values were not routinely tested every time in PUMCH), several SOFA scores, and total heparin dosage, as shown in Table 2. We used the last aPTT measurement before heparin treatment as the initial aPTT value, and the laboratory tests and SOFA scores were those closest to the initial heparin injection time.

XSL•FO
RenderX

Table 2. Clinical features of interest chosen from the 2 data sets.

Features	Multiparameter Intelligent Monitoring in Intensive Care III data set	Peking Union Medical College Hospital data set
Demographic data		
	Gender	Gender
	Ethnicity	—[a]
	Admission type	—
	Age	Age
	Weight	Weight
Laboratory tests		
	Initial aPTT[b] value	Initial aPTT value
	Creatinine value	Creatinine value
	AST/ALT[c] ratio	Alanine aminotransferase value
SOFA[d] scores		
	Coagulation SOFA score	Coagulation SOFA score
	Liver SOFA score	Liver SOFA score
	Cardiovascular SOFA score	Cardiovascular SOFA score
	Renal SOFA score	Renal SOFA score
Medication	Total heparin dosage	Total heparin dosage

[a]Not available.

[b]aPTT: activated partial thromboplastin time.

[c]AST/ALT: aspartate aminotransferase/alanine aminotransferase.

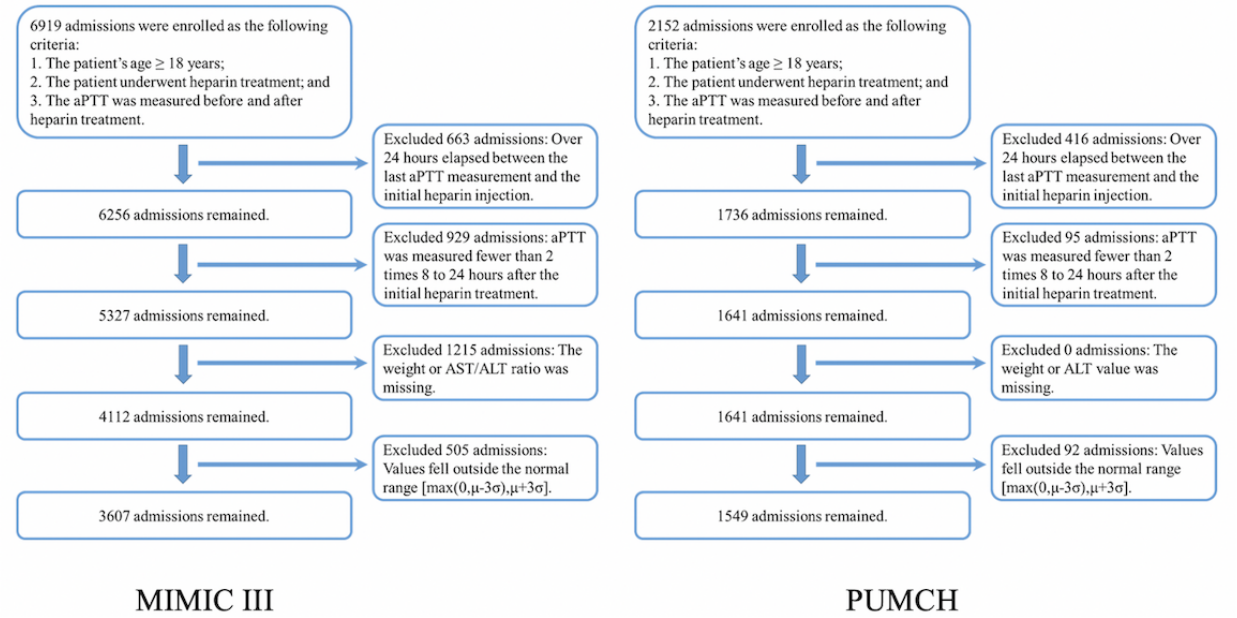
[d]SOFA: sequential organ failure assessment.

Patient Inclusion Criteria

The enrollment criteria were as follows: (1) patient's age ≥18 years, (2) patient underwent heparin treatment, and (3) the aPTT value was measured before and after heparin treatment. Based on the above criteria, we initially collected 6919 patient records from the MIMIC III database and 2152 patient records from the PUMCH database. For the MIMIC III database, we first removed some patient records whose aPTT data were unavailable. We then removed records with missing values for weight or the ratio of AST/ALT. Next, we removed patient records with values of continuous features outside the normal ranges, including weight, initial aPTT value, creatinine value, AST/ALT ratio, and total heparin dosage. According to the statistical definition of outliers [15], we calculated the mean value μ and standard deviation σ of these features; the normal range includes values from $\max(0, \mu-3\sigma)$ to $\mu+3\sigma$. We therefore removed the individual records outside of this range. The normal ranges and number of outliers for each feature are listed in Multimedia Appendix 1. After this process, 3607 patient records remained. For the PUMCH database, after a similar data selection process (according to different recording methods such as the replacement of the AST/ALT ratio by the ALT value in the PUMCH database), we collected 1549 patient records as the study population of PUMCH. The details of the selection process are shown in Figure 2.

XSL•FO
RenderX

Figure 2. Study cohort selection workflow based on the inclusion and exclusion criteria. ALT: alanine aminotransferase; aPTT: activated partial thromboplastin time; AST: aspartate aminotransferase; MIMIC III: Multiparameter Intelligent Monitoring in Intensive Care III; PUMCH: Peking Union Medical College Hospital.



Data Preprocessing

SOFA scores were missing in both data sets, and the 3 nearest neighbors algorithms were used to substitute the missing value with the mean of the 3 neighbors that was closest in terms of the Euclidean distance. The missing data imputation results are listed in Multimedia Appendix 2. We applied a one-hot encoder for the gender, ethnicity, and admission type features since these are categorical features. To avoid the occurrence of feature values that are too large for model training, we applied min-max normalization to all the continuous numerical features using the following formula:



where x denotes the values of all records for a fixed feature.

Model Training

As validated in a previous study [11], a shallow neural network model works best from among several machine learning models for the heparin outcome prediction task. In this study, we used a fully connected shallow neural network model [16,17] to predict the therapeutic effect (subtherapeutic, normal therapeutic, or supratherapeutic) of patients after 8 hours of heparin treatment. This is a ternary classification task. Our artificial neural network consists of an input layer, an output layer, and 3 hidden layers. It is fully connected, which means that each neuron receives input from all the neurons in the previous layer and passes on the results to all the neurons in the next layer. The rectified linear unit function [18] was used as the activation function in order to increase the nonlinearity and improve the model efficiency. The number of neurons in the hidden layers was set at 32/64/24. The models for MIMIC III and PUMCH data sets are basically the same, except for slight differences in the input. We trained the model for 5000 epochs with a learning rate of 0.015. The fully connected neural network was built using TensorFlow (version 1.12.0). Each data set was divided into training (80%) and test (20%) sets, with the proportion of subtherapeutic records, normal therapeutic records, and supratherapeutic records being maintained. To validate the predictive performance of our model, 5-fold cross-validation was used on each data set.

Heparin Dosage Recommendation

By using the neural network model described above, we can calculate the probability that a record belongs to each of the 3 categories (subtherapeutic, normal therapeutic, and supratherapeutic) with the softmax function [19]. For a subtherapeutic or supratherapeutic patient record x and a possible dosage α, we change the actual total heparin dosage of x to α and defined $M(x,\alpha)$ to be the normal therapeutic probability after this changing. Then, we calculated $M(x,\alpha)$ by the softmax function shown below:



where r, s, and t are the outputs (a larger number means a higher probability of belonging to this category) of the neural network model for supratherapeutic, normal therapeutic, and supratherapeutic, respectively, after this changing.

For this subtherapeutic or supratherapeutic patient record x, we provided a recommended total heparin dosage $\sigma(x)$ for x, which maximized the normal therapeutic probability. More specifically, we computed the recommended dosage using the following formula:



where O denotes the set of all possible dosages. We believe that our recommended dosage is reasonable and that our model may improve the clinician's judgement.

XSL•FO
RenderX

Model Evaluation

Model Performance Evaluation

The following measures, that is, precision, recall, F1 score, accuracy, and the kappa coefficient, were used to evaluate the capability of our ternary classification model [20,21]. Because the microaveraged precision, recall, and F1 score are all equal to accuracy, we only computed the accuracy, kappa coefficient, macroaveraged precision, recall, and F1 score to gauge the classification performance. For details, see the confusion matrix in Table 3 and the formulas below.

Table 3. The confusion matrix of our prediction.[a]

Actual label	Prediction label		
	Subtherapeutic	Normal therapeutic	Supratherapeutic
Subtherapeutic	A	B	C
Normal therapeutic	D	E	F
Supratherapeutic	G	H	I

[a]This table was used to gauge the classification performance, as shown in the formulas.



Feature Importance Evaluation

To evaluate the importance of each feature in our shallow neural network model, we removed 1 feature at a time and then calculated the decline in the model accuracy. More specifically, the importance E_i of the i-th feature F_i can be calculated by the formula $E_i = A - A_i$, where is the model accuracy when all features are used and A_i is the model accuracy after removing feature F_i. The features with higher E_i values are more important to the model. We also used 5-fold cross-validation on each data set to improve the stability of the calculation.

Recommendation Result Evaluation

The recommendation dosage result was compared with the actual decision made by the clinicians in the retrospective database, and the corresponding dosage response in the clinical scenario was also evaluated to validate whether the actual dosage setting was optimal. When the actual dosage setting was subtherapeutic and our recommended dosage was higher or it was supratherapeutic and our recommended dosage was lower, then we believe our recommended dosage was reasonable. For example, for a patient with predicted subtherapeutic aPTT outcomes after 10,000 IUs of heparin treatment, if a model recommended that dosage is greater than 10,000 IUs, we considered the recommended dosage to be reasonable; if the model recommended dosage is less than 10,000 IUs, we considered the recommended dosage to be unreasonable.

Results

Patient Characteristics

Patient characteristics according to the therapeutic outcome after heparin injection in MIMIC III and PUMCH are shown in Table 4. Among the 3607 MIMIC III patients, 1718 (47.6%) showed aPPT values within the subtherapeutic range, 1266 (35.1%) had values within the normal therapeutic range, and 623 (17.3%) patients had values within the supratherapeutic range. Among the 1549 PUMCH patients, 434 (28.1%) had measured aPPT values within the subtherapeutic range, 754 (48.7%) had values within the normal therapeutic range, and 361 (23.3%) had values within the supratherapeutic range. For numeric features, the feature density distribution was generated to clarify whether the values were scattered or centered (Multimedia Appendix 3).

XSL•FO
RenderX

Table 4. Patient characteristics and selected features according to the therapeutic outcome after heparin injection.

Data set, patient characteristics	Therapeutic range category		
	Subtherapeutic	Normal therapeutic	Supratherapeutic
Multiparameter Intelligent Monitoring in Intensive Care III database (n=3607)	1718	1266	623
Age (years), mean (SD)	65.4 (14.2)	67.8 (15.0)	70.6 (13.9)
Weight (kg), mean (SD)	84.5 (21.8)	81.3 (21.3)	80.5 (21.8)
Gender, n (%)			
Male	1068 (62.2)	750 (59.2)	320 (51.4)
Female	650 (37.8)	516 (40.8)	303 (48.6)
Ethnicity, n (%)			
White	1223 (71.2)	928 (73.3)	441 (70.8)
Black	98 (5.7)	111 (8.8)	71 (11.4)
Latin	42 (2.4)	21 (1.7)	16 (2.6)
Asian	24 (1.4)	25 (2.0)	25 (4.0)
Others	331 (19.3)	181 (14.3)	70 (11.2)
Admission type, n (%)			
Elective	189 (11.0)	69 (5.5)	22 (3.5)
Emergency	1474 (85.8)	1152 (91.0)	591 (94.9)
Urgent	55 (3.2)	45 (3.6)	10 (1.6)
Initial aPTT[a] value (s), mean (SD)	39.5 (22.0)	45.5 (26.8)	40.4 (21.2)
Creatinine value (mg/dL), mean (SD)	1.4 (1.0)	1.5 (1.1)	1.7 (1.2)
AST/ALT[b] ratio, mean (SD)	1.6 (1.1)	1.7 (1.2)	1.6 (1.0)
Coagulation SOFA[c] score, mean (SD)	0.5 (0.8)	0.4 (0.7)	0.4 (0.7)
Liver SOFA score, mean (SD)	0.4 (0.8)	0.4 (0.7)	0.4 (0.8)
Cardiovascular SOFA score, mean (SD)	1.5 (1.2)	1.5 (1.3)	1.7 (1.3)
Renal SOFA score, mean (SD)	0.8 (1.1)	1.1 (1.2)	1.3 (1.2)
Total heparin dosage (IUs), mean (SD)	8449.7 (6773.0)	11299.9 (7550.8)	12667.3 (6932.3)
Peking Union Medical College Hospital (n=1549)	434	754	361
Age (years), mean (SD)	55.1 (15.9)	57.8 (15.5)	60.9 (14.9)
Weight (kg), mean (SD)	68.1 (12.6)	67.3 (12.2)	66.5 (12.4)
Gender, n (%)			
Male	256 (59.0)	453 (60.1)	223 (61.8)
Female	178 (41.0)	301 (39.9)	138 (38.2)
Initial aPTT value (s), mean (SD)	29.5 (5.5)	34.6 (5.8)	39.4 (9.2)
Creatinine value (μmol/L), mean (SD)	107.4 (73.7)	117.2 (78.7)	128.0 (86.3)
Alanine aminotransferase value (unit/L), mean (SD)	45.1 (91.7)	47.1 (104.5)	55.7 (141.9)
Coagulation SOFA score, mean (SD)	0.9 (0.9)	1.0 (1.0)	1.1 (1.0)
Liver SOFA score, mean (SD)	0.6 (0.8)	0.7 (0.9)	0.8 (1.0)
Cardiovascular SOFA score, mean (SD)	2.9 (1.6)	3.0 (1.5)	3.1 (1.5)
Renal SOFA score, mean (SD)	0.5 (0.9)	0.6 (1.0)	0.8 (1.0)
Total heparin dosage (IU/kg), mean (SD)	7.5 (4.3)	7.8 (4.4)	9.6 (5.6)

[a]aPTT: activated partial thromboplastin time.

[b]AST/ALT: aspartate aminotransferase/alanine aminotransferase.

XSL•FO
RenderX

[c]SOFA: sequential organ failure assessment.

Model Performance

We divided each data set into a training (80%) set and test (20%) set. The proportions of each therapeutic level after dividing the data sets are listed in Multimedia Appendix 4. The shallow neural network model was trained and tested on a retrospective cohort of MIMIC III and PUMCH. The model performance results are listed in Table 5. For both data sets, our model achieved an accuracy of over 0.89, a kappa coefficient of over 0.82, and a macroaveraged F1 score of over 0.88. The results reflect the stable performance of the fully connected shallow neural network model and suggest that our model has generalizability for different databases.

Table 5. Model performance on the 2 data sets.

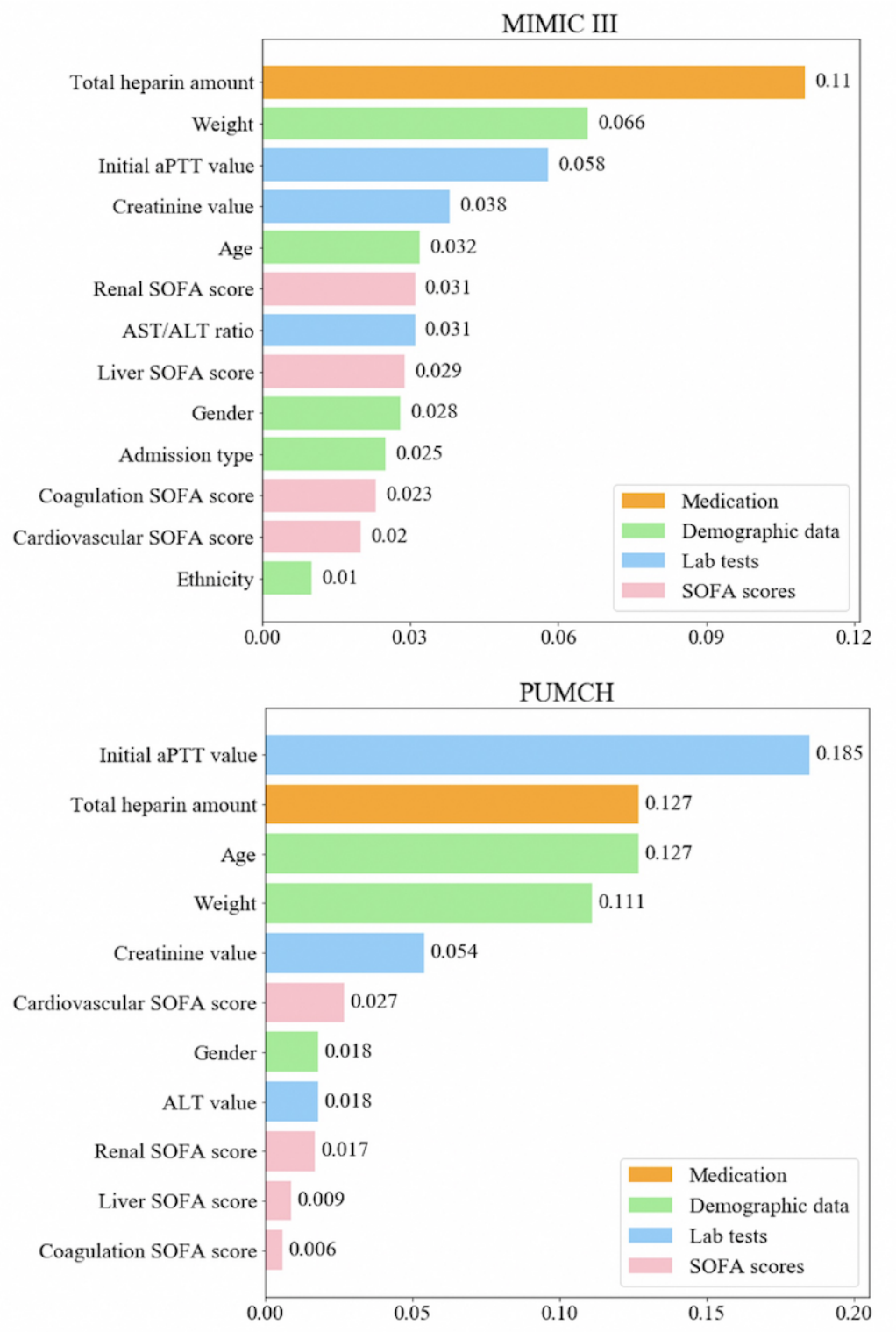
Parameters	Multiparameter Intelligent Monitoring in Intensive Care III data set	Peking Union Medical College Hospital data set
Accuracy	0.891	0.926
Kappa coefficient	0.823	0.882
Macroaveraged precision	0.890	0.931
Macroaveraged recall	0.884	0.920
Macroaveraged F1 score	0.887	0.925

Feature Importance

Figure 3 illustrates the importance of each feature in the development of our predictive model. The features are colored group-wise according to the categories listed in Table 2. A higher value indicates greater significance of the feature in the model. From the feature importance evaluation results, we conclude that the key features are basically the same in the 2 data sets but that their rankings differ slightly. For the MIMIC III data set, the total heparin dosage is the most influential model factor; weight, initial aPTT value, creatinine value, and age are key features and contributed substantially to the model. For the PUMCH data set, the highest impact feature is the initial aPTT value; total heparin dosage, age, weight, and creatinine are also important features for construction of the model. Ethnicity, SOFA scores, and gender have relatively small effects on the model in both data sets. Overall, the feature contribution analysis results are relatively consistent with clinical experience.

XSL•FO
RenderX

Figure 3. Feature importance. ALT: alanine aminotransferase; aPTT: activated partial thromboplastin time; AST: aspartate aminotransferase; MIMIC III: Multiparameter Intelligent Monitoring in Intensive Care III; PUMCH: Peking Union Medical College Hospital; SOFA: sequential organ failure assessment.



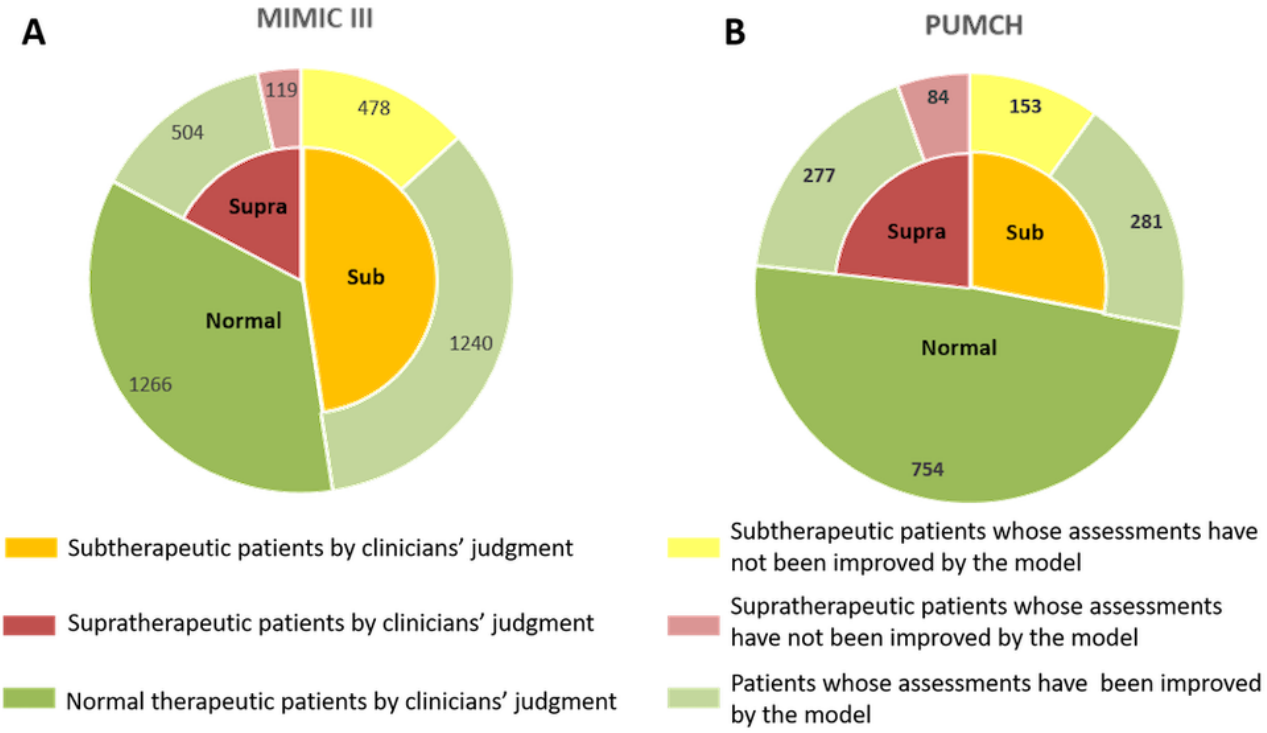
Evaluation Results of Recommended Heparin Dosage

We compared our recommended dosage with the actual dosage for subtherapeutic samples (MIMIC III: n=1718; PUMCH: n=434) and supratherapeutic samples (MIMIC III: n=623; PUMCH: n=361). For 72.2 % (1240/1718) of the subtherapeutic samples in MIMIC III and 64.7% (281/434) of the subtherapeutic samples in PUMCH, our model recommended a higher heparin dosage than the clinicians did. Moreover, for 80.9% (504/623) of the supratherapeutic samples in MIMIC III and 76.7% (277/361) of the supratherapeutic samples in PUMCH, our model recommended a lower heparin dosage than the clinicians did. In Figure 4, the solid yellow, green, and red in the inner circles represent the subtherapeutic, normal, and supratherapeutic samples, respectively. The same shaded colors

XSL•FO
RenderX

in the outer ring indicate that the recommended therapeutic state matches the inner circle, whereas fine green bars in the outer ring indicate the recommendation to increase the heparin dosage in subtherapeutic samples and to decrease the heparin dosage in supratherapeutic samples. The model recommendations may improve heparin treatment outcomes and may effectively reduce the time to optimal dosage in the clinical setting.

Figure 4. Recommend heparin dosage results in the MIMIC III (A) and PUMCH (B) data sets. MIMIC III: Multiparameter Intelligent Monitoring in Intensive Care III; PUMCH: Peking Union Medical College Hospital.



Discussion

Principal Results

In this study, we used a validated shallow neural network model for the clinical scenario of UFH infusion, which is commonly applied in the treatment and prevention of venous thromboembolism. To our knowledge, our study is the first practical validation of a machine learning–based model in the area of medication dosing optimization. We demonstrated the feasibility and efficiency of the proposed model in MIMIC III and in a local critical care database in a Chinese tertiary teaching hospital. Based on the calculated probabilities of individual circumstances, we proposed a UFH dosage recommendation for each record, and the comparison results show that the predicted recommended dosage can satisfactorily match the clinical practice.

Comparison With Prior Work

As described in a previous study by Ghassemi using the MIMIC II database [14], a large variation appears in the initial dosing and the corresponding aPTT response, which suggests an underlying discrepancy in both the interprovider practice and patient factors. In the clinical scenario, to avoid this deviation and subsequent fluctuations and treatment-related risks, in our local practice, instead of empirically setting an initial dosage, a step-by-step pattern was adopted. The applicable dosing level was determined according to a series of continuously monitored aPTT results. Therefore, in this study, we used the steady state aPTT value after 8 h of heparin treatment instead of a single value. The results show that the machine learning–based model can effectively predict the aPTT response after the initial dosing and in a step-by-step pattern, which can contribute to decreasing the duration of the therapeutic regime and avoiding treatment-related risks. In our previous work, we demonstrated that the shallow neural network algorithm performed best compared to algorithms such as extreme gradient boosting, adaptive boosting, and support vector machine [11]. Based on our local clinical database and modified treatment pattern, we validated our previously developed model and demonstrated the applicability of this machine learning–based algorithm for UFH treatment. Despite having trained the machine learning models in our previous study on 2 public data sets, which mainly come from the well-known MIMIC and electronic ICU databases, they may not be generalizable to other institutions and populations. Nevertheless, our experience has shown that in a specific clinical scenario, the models can be smoothly migrated to a new data set after retraining, which reflects the flexibility and simplicity of machine learning algorithms. It is worth noting that validation on different data sets demonstrates the generalizability of the model, but the model will not necessarily have the same set of coefficients for each data set.

The interpretability problem remains an issue in the application of machine learning algorithms to the clinical setting. Interpretability can contribute to a physician making decisions based on numerous clinical variables rather than simply providing a prediction or description [22]. In this study, we

XSL•FO
RenderX

calculated the importance of each feature in the shallow neural network model. As shown in Figure 3, the total heparin dosage, weight, initial aPTT value, creatinine value, and age are the 5 most important features for both data sets, although their specific ordering differs between data sets. In addition, as shown in Table 4, the average total heparin dosage increased successively from the subtherapeutic patients to the normal therapeutic patients and then to the supratherapeutic patients. It is reasonable that older age, lesser weight, and higher creatinine level tend to lead to supratherapeutic dosing. Conversely, younger age, higher weight, and lower creatinine level tend to lead to subtherapeutic dosing. This is consistent with clinical experience and can provide some practical support for clinicians. The feature importance analysis not only describes how the model works to a certain extent but can also prompt clinicians to pay more attention to important clinical variables. Although artificial neural network models are often regarded as black box models, we believe that the analysis and interpretation of these models will help understanding the model.

For each subtherapeutic or supratherapeutic patient, we recommended a total heparin dosage through the model. It is unreasonable to evaluate the recommended dosage with the model itself; therefore, we evaluated the rationality of the recommended dosage by comparing with the actual dosage given by the clinicians. It is noteworthy that in the MIMIC III data set, since the total heparin dosage is the most important feature, it contributed most to reasonable heparin dosage recommendation. In contrast, for the PUMCH data set, since the initial aPTT value is the most important feature and the total heparin dosage is relatively less important, the evaluation results of reasonable recommended heparin dosage are slightly low, as shown in Figure 4. Furthermore, the recommendation dosage results were initially evaluated in this study. For example, recommending increased dosage to subtherapeutic patients is evaluated as improved outcome; however, the recommended dosage may not definitely lead to a normal heparin treatment outcome—it may also lead to a supratherapeutic condition. We will further improve the evaluation method in our future studies.

Limitations

This study has several limitations. First, our study was limited by its retrospective nature and the sources of the data. The performance of this machine learning–based model should be validated in clinical practice in order to provide valuable suggestions for treatment. Therefore, in addition to model optimization and cross-validation using more clinical databases, a well-designed, prospective, crossover clinical study should be performed. Second, our results were limited by the size of the populations in the clinical databases; a larger cohort might contribute to more accurate predictions and more precise recommendations for the steady dosing level. Considering the similarity of the data and logical structure, other clinical scenarios such as therapeutic drug monitoring, homeostasis balancing, and blood transfusion control could be appropriate applications of this model. Third, as described above, considering the different time span, treatment pattern, and patient characteristics of the MIMIC III and PUMCH clinical databases, such as 2 feature sets are not completely consistent, the optimal dosing level and the corresponding aPTT results differ in the 2 data sets. Nevertheless, these disparities did not cause the dose-effect relationship to differ in either the clinical practice or the machine learning algorithm, as demonstrated by the model performance and feature importance results. Although input features need to be adjusted or preprocessed when applied to the other data set, we still regarded that the prediction model and recommendation method provide a machine learning solution when applied to heparin outcome prediction and decision-making clinical scenarios. Furthermore, from the perspective of model development, our fully connected shallow neural network model is currently a static model that does not make use of dynamic time series data. In addition, we have not incorporated other available relevant features into the model, such as medical history, comorbidities, surgery history, and intake of medications other than heparin. In the future, with the goal of achieving better predictions of the outcome of heparin treatment and recommending more reasonable heparin dosages, we will build a dynamic model such as a recurrent neural network or a long short-term memory model and incorporate more features.

Conclusions

Based on the machine learning model trained and validated in our previous work, this study aimed to further validate the model and its shallow neural network in a local clinical setting. We found that the data-driven machine learning method could be used effectively in the clinical scenario of UFH treatment with a step-by-step dosage pattern. The results provide support for predicting UFH treatment outcomes and recommending optimal UFH dosing to clinicians. We also evaluated the importance of each model feature to aid in the interpretation and understanding of the machine learning model.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (project 2018YFC0116903 and 2018YFC0116905), China Health Information and Health Care Big Data Association Severe Infection Analgesia and Sedation Big Data Special Fund (Z-2019-1-001), and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (project 2016-I2M-2-004).

Authors' Contributions

YL and WZ are the corresponding authors and take responsibility for the integrity of the whole work. DL, JG, NH, and HW contributed equally as co-first authors. DL, LS, and YL were responsible for study conception and design. HJ and WZ were responsible for data access and privacy management. JG, CL, QW, and HJ were responsible for data cleaning and algorithm

XSL•FO
RenderX

implementation. LS, DL, and HW were responsible for data analysis and explanation of results. DL, JG, and NH drafted the manuscript. All authors revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Normal ranges and number of outliers for each feature of clinical interest.
[DOCX File , 14 KB - jmir_v23i5e27118_app1.docx]

Multimedia Appendix 2
Missing data imputation results.
[DOCX File , 14 KB - jmir_v23i5e27118_app2.docx]

Multimedia Appendix 3
Density functions of different features.
[DOCX File , 12563 KB - jmir_v23i5e27118_app3.docx]

Multimedia Appendix 4
Proportions of each therapeutic level after dividing the data sets.
[DOCX File , 14 KB - jmir_v23i5e27118_app4.docx]

References

1. Hannes K, Leys M, Vermeire E, Aertgeerts B, Buntinx F, Depoorter A. Implementing evidence-based medicine in general practice: a focus group based study. BMC Fam Pract 2005 Sep 09;6:37 [FREE Full text] [doi: 10.1186/1471-2296-6-37] [Medline: 16153300]
2. Kuzniewicz MW, Vasilevskis EE, Lane R, Dean ML, Trivedi NG, Rennie DJ, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008 Jun;133(6):1319-1327. [doi: 10.1378/chest.07-3061] [Medline: 18403657]
3. Moffatt R. Book Review: Making Use of Guidelines in Clinical Practice. SAGE Publications Sage UK: London, England; 2001.
4. Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, ACTERREA Research Group. Big data and targeted machine learning in action to assist medical decision in the ICU. Anaesth Crit Care Pain Med 2019 Aug;38(4):377-384. [doi: 10.1016/j.accpm.2018.09.008] [Medline: 30339893]
5. 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]
6. Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. J Med Internet Res 2020 Jul 20;22(7):e18477 [FREE Full text] [doi: 10.2196/18477] [Medline: 32706670]
7. Melloni C, Alexander KP, Chen AY, Newby LK, Roe MT, Allen LaPointe NM, CRUSADE Investigators. Unfractionated heparin dosing and risk of major bleeding in non-ST-segment elevation acute coronary syndromes. Am Heart J 2008 Aug;156(2):209-215. [doi: 10.1016/j.ahj.2008.03.023] [Medline: 18657648]
8. Cruickshank MK, Levine MN, Hirsh J, Roberts R, Siguenza M. A standard heparin nomogram for the management of heparin therapy. Arch Intern Med 1991 Feb;151(2):333-337. [Medline: 1789820]
9. Hirsh J, Anand SS, Halperin JL, Fuster V, American Heart Association. AHA Scientific Statement: Guide to anticoagulant therapy: heparin: a statement for healthcare professionals from the American Heart Association. Arterioscler Thromb Vasc Biol 2001 Jul;21(7):E9-E9. [doi: 10.1161/hq0701.093520] [Medline: 11451763]
10. Lin R, Stanley MD, Ghassemi MM, Nemati S. A Deep Deterministic Policy Gradient Approach to Medication Dosing and Surveillance in the ICU. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:4927-4931 [FREE Full text] [doi: 10.1109/EMBC.2018.8513203] [Medline: 30441448]
11. Su L, Liu C, Li D, He J, Zheng F, Jiang H, et al. Toward Optimal Heparin Dosing by Comparing Multiple Machine Learning Methods: Retrospective Study. JMIR Med Inform 2020 Jun 22;8(6):e17648. [doi: 10.2196/17648] [Medline: 32568089]
12. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
13. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]

XSL•FO
RenderX

14. Ghassemi MM, Richter SE, Eche IM, Chen TW, Danziger J, Celi LA. A data-driven approach to optimized medication dosing: a focus on heparin. Intensive Care Med 2014 Sep;40(9):1332-1339 [FREE Full text] [doi: 10.1007/s00134-014-3406-5] [Medline: 25091788]
15. Cousineau D, Chartier S. Outliers detection and treatment: a review. Int J Psy Res 2010 Jun 30;3(1):58-67. [doi: 10.21500/20112084.844]
16. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 1982 Apr;79(8):2554-2558 [FREE Full text] [doi: 10.1073/pnas.79.8.2554] [Medline: 6953413]
17. Anderson JA, Silverstein JW, Ritz SA, Jones RS. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review 1977 Sep;84(5):413-451. [doi: 10.1037/0033-295X.84.5.413]
18. Glorot X, Yoshua B, Antoine B. Deep sparse rectifier neural networks. AISTATS. 2011. URL: http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf [accessed 2021-04-27]
19. Softmax function. Wikipedia. URL: https://en.wikipedia.org/wiki/Softmax_function [accessed 2021-04-21]
20. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management 2009 Jul;45(4):427-437. [doi: 10.1016/j.ipm.2009.03.002]
21. Cohen's kappa coefficient. Wikipedia. URL: https://en.wikipedia.org/wiki/Cohen%27s_kappa [accessed 2021-04-21]
22. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM computing surveys (CSUR) 2018;51(5):1-42. [doi: 10.1145/3236009]

Abbreviations

ALT: alanine aminotransferase
aPTT: activated partial thromboplastin time
AST: aspartate aminotransferase
ICU: intensive care unit
MIMIC III: Multiparameter Intelligent Monitoring in Intensive Care III
PUMCH: Peking Union Medical College Hospital
SOFA: sequential organ failure assessment
UFH: unfractionated heparin

Edited by R Kukafka; submitted 12.01.21; peer-reviewed by J Li, B Qian, L Wang; comments to author 29.01.21; revised version received 07.03.21; accepted 24.03.21; published 20.05.21.

Please cite as:
Li D, Gao J, Hong N, Wang H, Su L, Liu C, He J, Jiang H, Wang Q, Long Y, Zhu W
A Clinical Prediction Model to Predict Heparin Treatment Outcomes and Provide Dosage Recommendations: Development and Validation Study
J Med Internet Res 2021;23(5):e27118
URL: https://www.jmir.org/2021/5/e27118
doi:10.2196/27118
PMID:34014171

©Dongkai Li, Jianwei Gao, Na Hong, Hao Wang, Longxiang Su, Chun Liu, Jie He, Huizhen Jiang, Qiang Wang, Yun Long, Weiguo Zhu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Automatic Classification of Screen Gaze and Dialogue in Doctor-Patient-Computer Interactions: Computational Ethnography Algorithm Development and Validation

Samar Helou[1], PhD; Victoria Abou-Khalil[2], PhD; Riccardo Iacobucci[3], PhD; Elie El Helou[4], MD, MSc; Ken Kiyono[5], PhD

[1]Global Center for Medical Engineering and Informatics, Osaka University, Osaka, Japan

[2]Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan

[3]Department of Urban Management, Graduate School of Engineering, Kyoto University, Kyoto, Japan

[4]Faculty of Medicine, Saint Joseph University, Beirut, Lebanon

[5]Graduate School of Engineering Science, Osaka University, Osaka, Japan

Corresponding Author:
Samar Helou, PhD
Global Center for Medical Engineering and Informatics
Osaka University
Osaka Prefecture, Suita, Yamadaoka 2-2
Osaka, 565-0871
Japan
Phone: 81 8056856848
Email: helou.samar@gmail.com

Abstract

Background: The study of doctor-patient-computer interactions is a key research area for examining doctor-patient relationships; however, studying these interactions is costly and obtrusive as researchers usually set up complex mechanisms or intrude on consultations to collect, then manually analyze the data.

Objective: We aimed to facilitate human-computer and human-human interaction research in clinics by providing a computational ethnography tool: an unobtrusive automatic classifier of screen gaze and dialogue combinations in doctor-patient-computer interactions.

Methods: The classifier's input is video taken by doctors using their computers' internal camera and microphone. By estimating the key points of the doctor's face and the presence of voice activity, we estimate the type of interaction that is taking place. The classification output of each video segment is 1 of 4 interaction classes: (1) screen gaze and dialogue, wherein the doctor is gazing at the computer screen while conversing with the patient; (2) dialogue, wherein the doctor is gazing away from the computer screen while conversing with the patient; (3) screen gaze, wherein the doctor is gazing at the computer screen without conversing with the patient; and (4) other, wherein no screen gaze or dialogue are detected. We evaluated the classifier using 30 minutes of video provided by 5 doctors simulating consultations in their clinics both in semi- and fully inclusive layouts.

Results: The classifier achieved an overall accuracy of 0.83, a performance similar to that of a human coder. Similar to the human coder, the classifier was more accurate in fully inclusive layouts than in semi-inclusive layouts.

Conclusions: The proposed classifier can be used by researchers, care providers, designers, medical educators, and others who are interested in exploring and answering questions related to screen gaze and dialogue in doctor-patient-computer interactions.

(J Med Internet Res 2021;23(5):e25218) doi:10.2196/25218

KEYWORDS

computational ethnography; patient-physician communication; doctor-patient-computer interaction; electronic medical records; pose estimation; gaze; voice activity; dialogue; clinic layout

XSL•FO
RenderX

Introduction

Background

Doctor-patient communication is a combination of verbal and nonverbal expressions and can affect patient satisfaction, adherence, disclosure, and outcomes [1-8]. Health communication researchers have examined various aspects of clinician-patient verbal interactions, such as the content of the clinician's speech and their voice tone [5,9] and the intent that an utterance has in communication [10]. Various nonverbal aspects have also been examined, such as facial expressions, eye contact, body posture, fluency [5,11], and the physical distance between clinicians and patients [12]. With the widespread adoption of electronic medical record systems, computers have become an integral part of clinics. As a result, the traditional 2-way doctor-patient relationship has been replaced by a triadic relationship among doctor, patient, and computer [13]. The use of electronic medical record systems during consultations has been shown to affect doctor-patient verbal [14] and nonverbal [2] communication, and consequently, doctor-patient relationships both positively and negatively [15,16]. Accordingly, the study of doctor-patient-computer interactions has become a key research area for examining doctor-patient relationships [17].

Doctor-Patient-Computer Interactions

Multiple studies, mainly in primary care settings, noted that doctor-patient communication is affected [18-25] and even shaped [26] by the use of computers during clinical encounters. The use of computers was shown to modify or amplify doctors' verbal and nonverbal behaviors [16,21,27-29] that are essential to avoid communication failures and to have effective doctor-patient communication [20]. Examples of negative verbal and nonverbal behaviors that could be amplified by the use of a computer include lack of eye contact, deficient active listening, avoidance, and interruption [30-32].

In addition to studying the effect of computer use on doctor-patient interactions, multiple studies [25,33] examined factors that affect the way these computers are used. Pearce et al [34] described the overarching styles and behaviors of doctors, patients, and computers by studying the orientation of the general practitioners' and patients' bodies as well as their conversations. Chan et al [35] found that doctors spent 50% less time using computers in examinations with psychological components than in examinations with no psychological components. Lanier et al [36] found that consultation content, physicians' gender and level of experience, and whether the consultation was new or a follow-up were modestly related to the way physicians used the computer in primary care settings.

Computational Ethnography Inside Clinics

Researchers studying doctor-patient-computer interactions need to identify which interactions are taking place during the consultations. To do so, researchers have used qualitative methods such as taking notes during live observations [31,37], conducting interviews [37,38], administering questionnaires [39], and sending unannounced standardized patients to collect information [40] and quantitative methods such as videotaping consultations and manually coding the videos [36,41,42] or setting up complex mechanisms for automatic data collection and analysis inside the clinics [43]. Methods that include direct observations are likely to generate more accurate data than clinician or patient reports; however, direct observations are costly in terms of time and human resources, may be obtrusive in a clinical environment, and may cause the participants to knowingly or unknowingly alter their behavior (because of the presence of an observer) [44]. Moreover, they present privacy and ethical concerns for patients and doctors such as concerns about data security and anonymization; changes to the research question that make it different from the one described in initial consent forms, and researchers' inability to take into account all nonpublic information or situations that will be accessed [45].

Given recent technological advancements, computational ethnography has been proposed as an alternative method for studying doctor-patient-computer interaction in depth. Computational ethnography was defined as a new family of methods for conducting human-computer interaction studies in health care settings by using "automated and less obtrusive (or unobtrusive) means for collecting in situ data reflective of real end users' actual, unaltered behaviors using a software system or a device in real-world settings [46]."

Recently, a number of tools that automate the measurement and analysis of specific behaviors in clinical settings were proposed and evaluated: Hart et al. [47] proposed and validated an automated video analysis tool to measure the synchrony and dominance in doctor-patient interactions by analyzing the cross-correlation of the kinetic energy and the frequency spectrum of their motion [47]. Gutstein et al reported developing a system that automatically learns the physician's gaze using their hand positioning [48] or body positioning and optical flow [49]. Weibel et al [43] introduced a solution that enables the capture of multimodal activity in clinical settings [43] to support computational ethnography studies in clinics. Their solution combined computer logging functionality, body motion tracking, audio detection, and eye tracking. By synchronizing data from these sensors, Weibel et al [43] were able to detect the person talking, whether the doctor is looking at the screen, the amount of gesturing, the cognitive load, information searching behavior, workflow interruptions, and the amount of computer activity; however, their solution had some limitations. First, the accuracy of the automatic classification was not reported. Second, they noted that the use of Kinect presents some limitations such as the need to set up the machine, Kinect's inability to reidentify a body once it re-enters the scene, and the occasional transfer of skeletal tracking from human to nonhuman objects. Third, to detect the person who was talking, a Dev-Audio Microcone [50] was used. This means that such a tool may not fit the needs of people looking for a cheap and portable solution with a known robustness level. In this case, recent advancements in pose and voice activity detection algorithms could address some of these limitations. For video consultations, Faucett et al [51] created ReflectLive, a tool that provides real-time feedback to clinicians about speaking contributions, interruptions, eye gaze, and face position. ReflectLive [51] uses an open-source library for audio analysis and a commercial Javascript-based computer-vision

XSL•FO
RenderX

face-tracking software for visual analysis. The real-time feedback provided by ReflectLive was evaluated in terms of its usefulness to the clinicians, but the feedback's accuracy was not reported.

Currently, there are few truly robust and unobtrusive computational ethnography tools for clinical settings, as most tools require researchers to add external artifacts into the clinical environment. Moreover, to our knowledge, none of the existing tools is freely available to the public. To enable human-computer and human-human interaction studies in clinical settings, there is a need for publicly available, robust, unobtrusive, and automated tools for detecting and classifying doctor-patient-computer interactions.

Aims

We aimed to provide a public, robust, unobtrusive, privacy-ensuring, and automated tool for detecting and classifying screen gaze and dialogue in doctor-patient-computer interactions. We chose to focus on screen gaze and dialogue due to recent advancements in machine learning that render the automatic and accurate estimation of pose and voice activity possible.

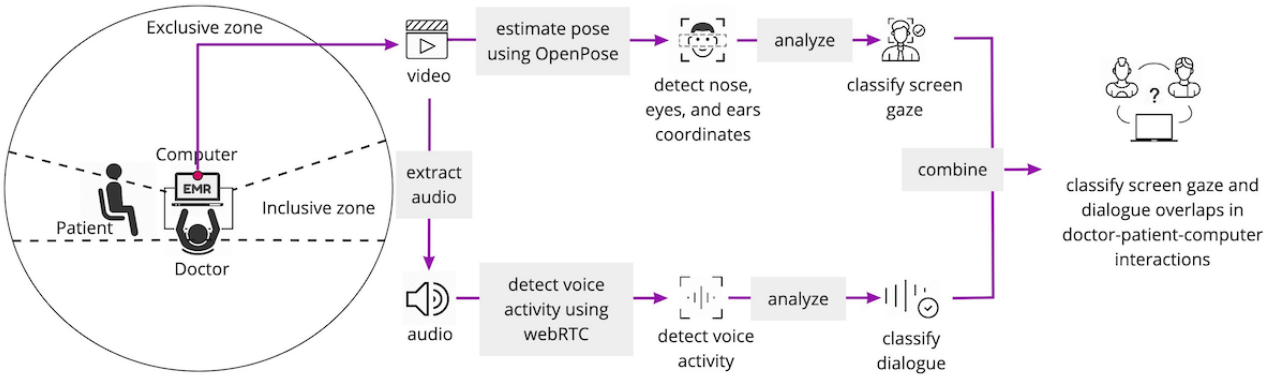
Methods

Overview

The purpose of the classifier (Figure 1) was to detect the following interactions: (1) screen gaze and dialogue: doctor gazing at the computer screen while having a conversation with the patient; (2) dialogue: doctor conversing with the patient while looking away from the computer screen, or (3) screen gaze: doctor gazing at the computer screen without conversing with the patient. Any other type of interaction in which the doctor and the patient were not having a conversation and the doctor is not gazing at their computer screen were considered out of scope.

The code of the proposed classifier is publicly available [52] and can be used by researchers, care providers, designers, medical educators, and others who are interested in exploring and answering questions related to screen gaze and dialogue combinations in doctor-patient-computer interactions.

Figure 1. Overview of the classification process. EMR: electronic medical record.



Screen Gaze Classifier

The purpose of the screen gaze classifier was to detect when the doctor's gaze was aimed at the computer screen. The input of the classifier was the video captured by the doctor's computer camera and the output was a binary classification: *no screen gaze* or *screen gaze*.

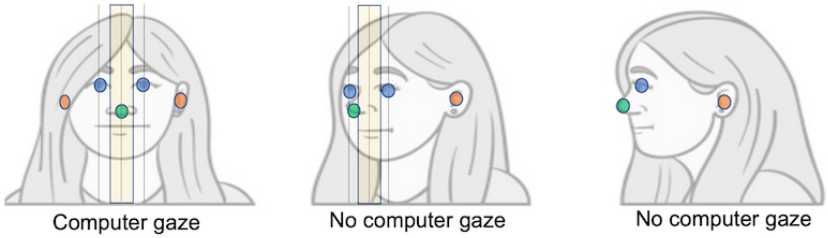
We used the pose estimation library OpenPose [53] as a tool to detect the coordinates of key points of the doctor's face. OpenPose is an open-source library that allows real-time multiperson key point detection for body, face, hands, and feet. We extracted the coordinates of the doctor's eyes ($x_{LeftEye}$, $y_{LeftEye}$), ($x_{RightEye}$, $y_{RightEye}$), ears ($x_{LeftEar}$, $y_{LeftEar}$), ($x_{RightEar}$, $y_{RightEar}$), and nose (x_{Nose}, y_{Nose}), and using the coordinates, we assumed that the doctor's gaze was targeting the computer screen if (1) the location of both the doctor's ears could be estimated, and (2) the doctor's nose was centered between the eyes. For the second condition, we allowed a tolerance equal to half the distance between the 2 eyes. We assessed these criteria (Figure 2) using the following equations for each frame in the video:



Then we assigned to each 0.5-second interval of video the most frequent classification of its corresponding frames. This results in a binary classification (no screen gaze, screen gaze) for each 0.5 seconds of video.

XSL•FO
RenderX

Figure 2. Classifying the doctor's computer screen gaze using face key point estimation.



Dialogue Classifier

The purpose of the dialogue classifier was to detect when the doctor and patient were engaging in conversation. The input of the classifier was the audio captured by the doctor's computer's microphone, and the output was a binary classification of the doctor-patient conversation: *no dialogue* or *dialogue*.

We used a library based on the webRTC voice activity detection engine (an open source project maintained by the Google WebRTC team [54]). The voice activity detection library allows the detection of voice activity in an audio file by processing audio segments and estimating the probability in each segment.

We set the length of each audio segment to 5 milliseconds, which we found to offer the best results through trial and error. We set the voice activity detection to its highest aggressiveness mode in order to increase the probability of filtering out nonspeech. We assigned to each 0.5-second interval of audio the most frequent classification of its corresponding segments. This results in a binary classification (no dialogue, dialogue) for each 0.5 seconds of audio.

Classifier of Screen Gaze and Dialogue Combinations

By combining the results of the screen gaze classifier and the dialogue classifier described above, we classify doctor-patient-computer interactions into 4 different classes (Table 1). The Screen Gaze and Dialogue (SG+D) class defines interactions wherein the doctor is gazing at the computer screen while conversing with the patient. The Dialogue (D) class defines interactions wherein the doctor is looking away from the computer screen and conversing with the patient, and the Screen Gaze (SG) class defines interactions where the doctor is gazing at the computer screen and not conversing with the patient. An interaction wherein the doctor is neither looking at the computer screen nor conversing with the patient is classified as Other. For each 0.5 seconds of video, the interactions classifier assigns 1 of the 4 classes.

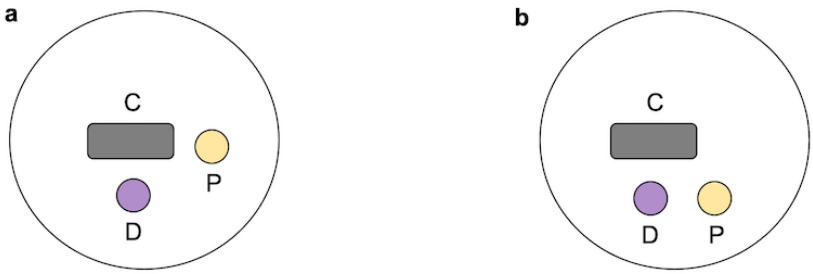
Table 1. Four classes of doctor-patient-computer interactions.

Components		Class	
Screen gaze	Dialogue	Doctor-Patient-Computer interaction	Label
Screen gaze	Dialogue	Screen Gaze + Dialogue	SG+D
No screen gaze	Dialogue	Dialogue	D
Screen gaze	No dialogue	Screen Gaze	SG
No screen gaze	No dialogue	Other	Other

Evaluation of the Classifier

We considered 2 clinical layouts in our evaluation: a semi-inclusive layout, where the patient is seated next to the computer desk, and a fully inclusive layout, where the patient is seated next to the doctor and facing the computer desk (Figure 3). The data that we used to evaluate our classifier consisted of 10 videos provided by 5 physicians. Each physician provided 2 videos—1 video simulating a consultation in a fully inclusive layout and 1 video simulating a consultation in a semi-inclusive layout. Each video was approximately 3 minutes long.

Figure 3. (a) Semi-inclusive and (b) fully inclusive layouts were considered in the evaluation. C: computer; D: doctor; P: patient.



XSL•FO
RenderX

For ground truth data, each video was initially annotated by a human coder. The coder assigned 1 of the 4 interaction classes to each 0.5 seconds of video. The coder reviewed the videos several times and refined the initial annotations until they were satisfied.

To evaluate the classifier, we compared the classifier's performance to that of a different human coder. This second coder was allowed to go through the video only once. This was to simulate a real-world scenario of video coding assigned to an external coder. The performance of the classifier and that of the second human coder were assessed in relation to the ground truth data (generated by the first coder).

The overall performance reflects the performance over the 10 videos including 5 videos in a semi-inclusive layout and 5 videos in a fully inclusive layout. The performances were assessed using an overall accuracy measure in addition to measures of precision, recall, F1 scores for each class. For each class, the support number (ie, the number of its occurrences in the ground truth data set) is reported. Weighted scores of precision, recall, and F1 scores, where the weight of a class is proportional to its support, were also measured. Difference in performance between the classifier and the human coder were assessed using 2-tailed independent *t* tests with *P* values <.05 considered statistically significant. We first report the overall performance, which reflects the performance over the 10 videos. Then, we separately report the performances over the 5 videos in a semi-inclusive layout and the 5 videos in a fully inclusive layout.

Results

Overall Performance

Table 2 shows the overall performances of the classifier and the human coder. The classifier showed a slightly lower overall accuracy than the coder (classifier: 0.83; human coder: 0.85); however, there was no significant difference between the accuracy of the classifier and that of the human coder (t_{18}=0.6, *P*=.55).

The F1 scores of both the classifier and the coder were better when classifying SG+D (classifier: 0.81; human coder: 0.81) and D (classifier: 0.89; human coder: 0.90) than that when classifying SG (classifier: 0.63; human coder: 0.55) and Other (classifier: 0.35; human coder: 0.36) interactions. Since the D class and the SG+D class were the most frequent interactions (D: 2415/3921, 62%; SG+D: 1189/3921, 30%), the overall accuracies mainly reflect performances for these 2 classes.

Confusion matrices for overall performance (Figure 4) show that the classifier and the coder had similar patterns. Both mistook SG+D for D and vice versa, SG for SG+D, and Other for D interactions. The main difference between the classifier and the coder is that the classifier tended to mistake D for Other interactions, whereas the coder tended not to.

Table 2. Overall performance of the classifier.

Classes	Classifier				Human coder				Support, n
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.83	—	—	—	0.85	—
SG+D[b]	0.79	0.82	0.81	—	0.78	0.84	0.81	—	1189
D[c]	0.92	0.86	0.89	—	0.91	0.90	0.90	—	2425
SG[d]	0.64	0.63	0.63	—	0.71	0.45	0.55	—	228
Other	0.24	0.67	0.35	—	0.35	0.38	0.36	—	79
Weighted score	0.85	0.83	0.84	—	0.85	0.85	0.84	—	—

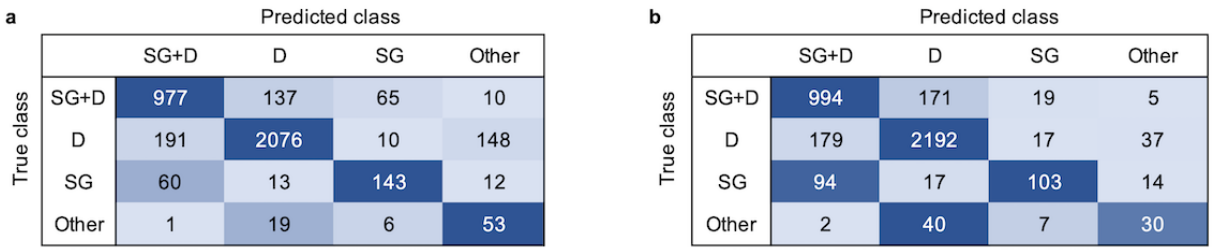
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 4. (a) Classifier and (b) human coder confusion matrices for overall performance. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



XSL•FO
RenderX

Performance in a Semi-inclusive Layout

Table 3 shows the performances of the classifier and human coder for a semi-inclusive layout. The classifier had a slightly lower accuracy than the coder (classifier: 0.80; human coder: 0.83); however, there was no significant difference between the accuracy of the classifier and that of the human coder in the semi-inclusive layout (t_8=1.04, P=0.32).

Both the classifier and the coder performed well when classifying D (classifier: F1 score 0.86; human coder: F1 score 0.88) and SG+D (classifier: 0.79; human coder: 0.82) interactions. The classifier had a slightly better F1 score than the coder when detecting SG (classifier: 0.47; human coder: 0.45), but both the classifier and the coder had low F1 scores when classifying Other interactions (classifier: 0.24; human coder: 0.21). The D and the SG+D classes had the most support (D: 1157/1958, 59%; SG+D: 702/1958, 36%) thus the overall accuracies mainly reflect performances for these 2 classes.

Confusion matrices of the classifier and the human coder (Figure 5) for a semi-inclusive layout show somewhat similar patterns for the classifier and the coder. Both mostly mistook D for SG+D and vice versa. The classifier tended to mainly mistake SG for SG+D, whereas the coder mistook SG for Other interactions as well. Finally, the coder exceedingly mistook Other for D.

Table 3. Performance of the classifier in a semi-inclusive layout.

Classes	Classifier				Human coder				Support, n
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.80	—	—	—	0.83	—
SG+D[b]	0.76	0.83	0.79	—	0.81	0.82	0.82	—	702
D[c]	0.92	0.80	0.86	—	0.88	0.88	0.88	—	1157
SG[d]	0.40	0.57	0.47	—	0.61	0.36	0.45	—	69
Other	0.16	0.50	0.24	—	0.17	0.30	0.21	—	30
Weighted score	0.83	0.80	0.81	—	0.84	0.83	0.83	—	—

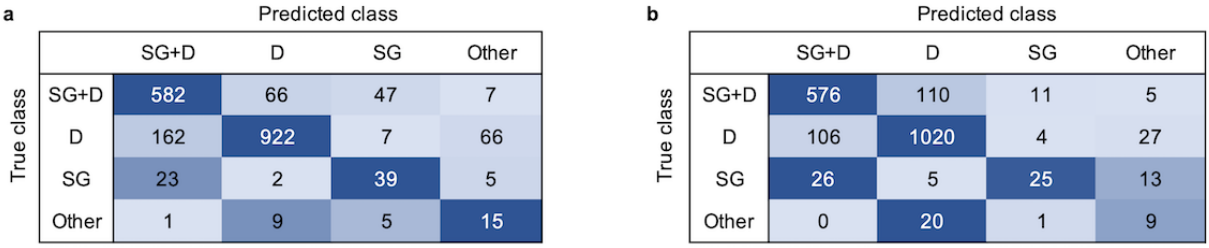
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 5. (a) Classifier and (b) human coder confusion matrices for semi-inclusive layout. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



Performance in a Fully Inclusive Layout

Table 4 shows the performances of the classifier and the human coder in a fully inclusive layout. The classifier and the coder showed similar accuracy (both equal to 0.86), and there was no significant difference between the accuracy of the classifier and that of the human coder for a fully inclusive layout (t_8=0.43, P=0.67).

The classifier and the coder had good F1 scores when classifying D (classifier: 0.92; human coder: 0.93) and SG+D (classifier: 0.83; human coder: 0.80). The classifier performed better than the coder for the SG class (classifier: 0.73; human coder: 0.59), but worse for Other interactions (classifier: 0.42; human coder: 0.52). The D and SG+D classes had the most support(D: 1268/1963, 65%; SG+D: 487/1963, 25%), thus the overall accuracy mainly reflects the performance for these 2 classes.

Confusion matrices of the classifier and the human coder for a fully inclusive layout (Figure 6) show similar patterns for the classifier and the coder. Both mistook SG+D for D, SG for SG+D, and Other interactions for D; however, the classifier tended to mostly mistake D for Other interactions, whereas the coder mostly mistook D for SG+D.

XSL•FO
RenderX

Table 4. Performance of the classifier in a fully inclusive layout.

Classes	Classifier				Human coder				Support
	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	
All	—[a]	—	—	0.86	—	—	—	0.86	—
SG+D[b]	0.86	0.81	0.83	—	0.75	0.86	0.80	—	487
D[c]	0.93	0.91	0.92	—	0.93	0.92	0.93	—	1268
SG[d]	0.83	0.65	0.73	—	0.74	0.49	0.59	—	159
Other	0.29	0.78	0.42	—	0.66	0.43	0.52	—	49
Weighted score	0.88	0.86	0.87	—	0.86	0.86	0.86	—	—

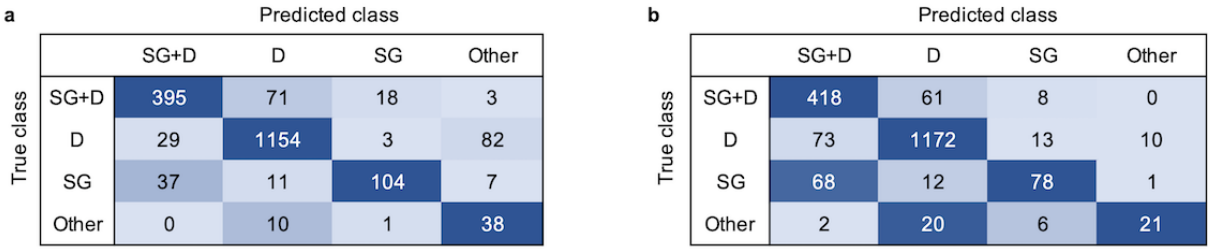
[a]Not calculated or not applicable.

[b]SG+D: Screen Gaze and Dialogue.

[c]D: Dialogue.

[d]SG: Screen Gaze.

Figure 6. (a) Classifier and (b) human coder confusion matrices for fully inclusive layout. D: Dialogue; SG: Screen Gaze; SG+D: Screen Gaze + Dialogue.



Transitions Between Doctor-Patient-Computer Interactions

We found that many errors in the classifications (for both the human coder and our classifier) were due to slight time inconsistencies during transitions. To confirm this hypothesis, we conducted an analysis of the transitions between the interactions. Table 5 reports the frequency of each transition in the ground truth data and shows that most transitions happen from D to SG+D and vice versa. We define a transition timing error as a temporal shift of 0.5 to 1 seconds between the ground truth and the classification. We only included transitions that were preceded and followed by a continuous type of interaction for at least 1.5 seconds. Table 6 reports the absolute and relative number of errors that can be attributed to early or late coding of transitions. We think these errors can be overlooked for any practical purpose.

XSL•FO
RenderX

Table 5. Transitions in ground truth data.

Transition	Semi-inclusive, n	Fully inclusive, n
From SG+D[a]		
...to D[b]	103	65
...to SG[c]	10	16
...to Other	1	0
From D		
...to SG+D	104	67
...to SG	7	15
...to Other	10	16
From SG		
...to SG+D	8	13
...to D	10	17
...to Other	0	2
From Other		
...to SG+D	1	1
...to D	10	14
...to SG	2	4

[a]SG+D: Screen Gaze and Dialogue.

[b]D: Dialogue.

[c]SG: Screen Gaze.

Table 6. Transition-related errors.

Layout	Classifier		Human coder	
	Total errors, n	Transition errors, n (%)	Total errors, n	Transition errors, n (%)
Semi-inclusive	400	65 (16.2)	329	73 (22.2)
Fully inclusive	272	45 (16.5)	274	61 (22.3)

Discussion

Principal Results

We developed an unobtrusive, inexpensive, and automatic classifier of screen gaze and dialogue combinations in doctor-patient-computer interactions. The classifier was evaluated in 2 clinical layouts, semi-inclusive and fully inclusive, and had a performance similar to that of a human coder with an overall accuracy of 0.83. The proposed classifier is unobtrusive since it does not require additional setup in the clinic and only requires that doctors record video using their computer's internal microphone and camera. The proposed classifier is an inexpensive solution since it is built using open-source tools and takes advantage of the internal camera and microphone built into most available computing devices. Finally, the video can be locally processed, thus reducing the risks of handling private and sensitive data off the clinic's premises, and ensuring that no collateral data are collected and used for purposes other than those initially consented to by the participants.

Both the classifier and the coder had better accuracies in a fully inclusive layout (both equal to 0.86) than in the semi-inclusive layout (classifier: 0.80; human coder: 0.83). The difference in performance can be attributed to the different postures that the doctor maintains when interacting in the 2 clinic layouts. In the fully inclusive layout, the doctor has to rotate their head a full 90 degrees away from the screen in order to gaze at the patient, whereas in semi-inclusive scenarios, the head rotation angle is smaller; therefore, it is easier to make distinctions between the interactions in a fully inclusive layout.

Both the classifier and the coder confused SG+D interactions and D interactions. Some instances occurred when classifying near transitions between interactions. Indeed, our analysis showed that 16.4% of the classifier's errors (110/672) and 22.2% of the human coder's errors (134/603) were early or delayed markings of transitions, which can be overlooked in practical use-cases. Unlike the human coder, the classifier tended to mistake D for Other interactions (ie, the doctor is neither looking at the screen nor conversing with the patient). This may be attributed to the fact that human coders overlook small moments of silence and regard them as response offsets that are due to turn-taking in the conversation [55] or lapses that are expected

XSL•FO
RenderX

in multiactivity settings [56,57], whereas our classifier classifies them as an absence of dialogue. Here, the classifier presents an advantage since it easily detects moments of silence that are usually overlooked by human coders. This kind of information may be useful to detect conversational dimensions such as hesitant speech [42]. However, for our purposes, this leads the classifier to overestimate the lack of verbal interaction. To counteract this, further rules are needed to identify which moments of silence are part of a conversation and which are not. These rules need to take into consideration the language and the content of the conversation [55,58], other activities that individuals are engaging in while conversing [56,57], and accompanying nonverbal behavior such as nodding and gaze [55].

As a collateral result, our experiment also confirmed the findings of previous work that highlighted the effect of the clinic layout on doctor-patient-computer interactions [59,60]. Fully inclusive videos contained more D (64% versus 59%) and SG interactions (8% versus 3.5%) and fewer SG+D interactions (25% versus 36%) than those in semi-inclusive videos. In a fully inclusive layout, the doctor has to choose whether to face the computer or their patient, whereas a semi-inclusive layout allows the doctors to maintain a conversation with their patients while looking at the computer screen.

Use Cases

Functionality

Because health communication researchers study various complex behaviors, the coding process is far more complex than a binary classification of screen gaze and dialogue over time. However, the proposed classifier could contribute to reducing the cost of future studies since gaze and dialogue are part of the behaviors that health communication researchers and medical informaticians are often interested in quantifying to examine the relationship between computer use, clinic and tool design, and physician-patient interactions [3,5,19,28,61-65]. Moreover, by detecting these behaviors over short intervals of time, more complex behaviors and patterns could be inferred as shown in our results on transitions between interactions. Therefore, the classifier can be directly used to monitor screen gaze and dialogue or extended and combined with other tools or processes to examine complex interactions in clinical settings.

For Researchers

The proposed tool can be used to study the effects of screen gaze and dialogue combinations in doctor-patient-computer interactions on quality of care and health outcomes. Currently, conducting this type of study would require (1) recording a video of the consultation, (2) transferring the video outside of the clinic, (3) manual coding of screen gaze and presence of dialogue in the video, and (4) assessing the specific outcome. This process can be facilitated by our classifier, which eliminates the need for the second and third steps.

The classifier can also be used in clinics to examine the effect of external factors on screen gaze and dialogue combinations in patient-doctor-computer interactions such as sociodemographic characteristics, clinic layout, and modifications of electronic medical record system's design. Non–self-reported large-scale studies of this kind would be nearly impossible to conduct using current methods and tools.

In addition, the gaze classifier can be used in conjunction with commonly used interaction analysis coding systems, such as the Roter interaction analysis system [42], that do not systematically account for nonverbal behaviors [66]. This would provide useful data for studies examining provider-patient communication in the presence of a computer.

For Practicing Physicians

Since the processing of the video can happen in real time, tools that allow physicians to *reflect in action* [67] can be created. The classifier can be used to create tools that allow physicians to conduct autoethnographies and reflect on their interactions with the patient, or on the role of technology in their care practice [68]. This would allow them to adapt their behavior and level of attention based on feedback. Similar concepts were proposed by Liu et al [69] and Faucett et al [51], who described tools that provide feedback to clinicians about their verbal and nonverbal communication behaviors during online consultations. Their studies highlighted the usefulness of summative [70] and real-time self-reflection tools and the need to design real-time feedback in a way that minimizes intrusiveness and ensure that it does not create extra distractions [51].

For Medical Educators and Students

Medical educators and students can use the classifier to teach and learn the best practices of doctor-patient interactions. The tool could be expanded to detect different interaction categories (eg, listening/ignoring; confronting/avoiding [71,72]) and used during practical learning sessions to provide students with formative feedback [69].

Limitations and Future Work

The first limitation of the classifier is its applicability to certain clinic layouts. Our evaluation explored 2 clinical scenarios: semi-inclusive and fully inclusive. We did not explore exclusive scenarios, even though these scenarios might be encountered in real-world clinical settings. With the proposed classifier, detecting the doctor's computer gaze would not be possible in a fully exclusive scenario since the classifier considers the doctor's head turn to estimate her gaze. We consider this limitation acceptable since inclusive scenarios are already commonplace [25], and we expect an increase in their prevalence to support technology-mediated information-sharing between clinicians and patients. Indeed, previous studies [31,73,74] have shown that clinicians use their computer screens as tools to share information with their patients; therefore, the doctor may turn the screen, along with the camera, toward the patient. If this interaction happens during an ongoing conversation, our tool may classify it as a dialogue between doctor and patient, but the fact that this doctor-patient interaction is mediated by the computer would not be highlighted. Further improvements are needed to detect scenarios where both clinician and patient are interacting with the computer at the same time.

Our work also assumes that clinicians use their computers during consultations, which is not always the case, especially in

XSL•FO
RenderX

secondary care settings. Moreover, the classifier is built and evaluated around the premise that the only people in the clinic are the doctor and the patient and that the only screen is the doctor's computer screen. However, it is possible that patients are accompanied by their family members, friends, or partners [75] and that multiple health care staff are involved in the care of 1 patient and present during the consultation. In addition, extra screens might be installed inside clinics to engage the patient in their care and offer them an easy and clear view of their data. These screens may affect the doctor's behavior in various ways; for example, they might use this screen as an explanation support tool while they converse with the patient or even as their main computer screen. Therefore, another limitation of this work is its nonapplicability in scenarios that include patient screens and stakeholders other than the doctor and patient.

Furthermore, the classifier does not allow us to identify the speaker's identity or the content of the doctor-patient dialogue. Therefore, the classifier does not currently support conversation analysis. To identify the speaker's identity, we would have to perform accurate speaker diarization, a hard goal to achieve especially using a single channel for audio recording and without prior training. Advancements in speaker diarization techniques may render this task feasible in the near future [76]. To identify the content of the dialogue, automatic speech recognition solutions can be used. Though automatic speech recognition solutions have become more robust in the last decade, the performance of automatic speech recognition engines remains limited when applied to conversational clinical speech [77]. Future work could explore the feasibility of automatic conversation analysis in doctor-patient-computer interactions through the application of novel speaker diarization and automatic speech recognition tools.

In other respects, although the direction of the head could be considered a proxy for the direction of attention, the head only communicates short-term attention [78,79]. Pearce et al classified physicians as unipolar, those who maintain the lower pole of their body facing the computer, or bipolar, those who repeatedly alternate the orientation of their lower pole between the computer and the patient [34]. Unipolar physicians experience situations where their body segments are not aligned, also referred to as body torque. Body torque communicates an instability of attention where the most strongly projected resolution involves the upper body getting realigned with the lower body. This means that the orientation of the torso communicates longer-term attention than head orientation, and the orientation of the legs communicates longer-term attention than torso orientation and head orientation [79]. Therefore, to examine the attention of a physician in a clinical scenario, we also need to examine the orientation of their torso and lower body. Future work can use the same pose estimation approach to monitor the direction of the clinician’s torso. This is possible because the shoulders and the torso of the clinician are usually visible to their computer’s camera; however, monitoring the lower part of the body would require setting up extra cameras in the clinic.

Finally, our classifier is model-driven as it derives its decisions from the explicit rules that we set. To be able to classify interactions that do not fit strictly into our specified rules, the classifier has to be driven by data or perhaps be a system that combines model-driven and data-driven logic. Our future work will include collecting and annotating more videos of consultations in order to create a data-driven classifier.

Conclusions

To facilitate human-computer and human-human interaction studies in clinical settings, we presented a computational ethnography tool—an automatic unobtrusive classifier of gaze and dialogue combinations in doctor-patient-computer interactions. The classifier only requires that the doctor record video using their computer’s internal camera and microphone. Our evaluation showed that the classifier's performance was similar to that of a human coder when classifying 3 combinations of screen gaze and dialogue in doctor-patient-computer interactions.

Acknowledgments

This research was supported by Japan Society for the Promotion of Science Kakenhi (grant number JP20K20244).We thank the 5 physicians who provided us with video data. We also thank our reviewers for their valuable comments and suggestions, which helped us improve this work.

Conflicts of Interest

None declared.

References

1. Hall JA. Affective and nonverbal aspects of the medical visit. In: The Medical Interview, Frontiers of Primary Care. New York, NY: Springer; 1995:495-503.
2. Robinson JD. Nonverbal communication and physician-patient interaction: review and new directions. In: Manusov V, Patterson ML, editors. The SAGE Handbook Of Nonverbal Communication. Thousand Oaks, CA: SAGE Publications, Inc; 2006:437-460.
3. Roter D, Frankel R, Hall J, Sluyter D. The expression of emotion through nonverbal behavior in medical visits. mechanisms and outcomes. J Gen Intern Med 2006 Jan;21 Suppl 1(1):S28-S34 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00306.x] [Medline: 16405706]

XSL•FO
RenderX

4. Beck RS, Daughtridge R, Sloane PD. Physician-patient communication in the primary care office: a systematic review. J Am Board Fam Pract 2002;15(1):25-38 [FREE Full text] [Medline: 11841136]
5. Griffith CH, Wilson JF, Langer S, Haist SA. House staff nonverbal communication skills and standardized patient satisfaction. J Gen Intern Med 2003 Mar;18(3):170-174 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.10506.x] [Medline: 12648247]
6. Schoenthaler A, Chaplin WF, Allegrante JP, Fernandez S, Diaz-Gloster M, Tobin JN, et al. Provider communication effects medication adherence in hypertensive African Americans. Patient Educ Couns 2009 May;75(2):185-191 [FREE Full text] [doi: 10.1016/j.pec.2008.09.018] [Medline: 19013740]
7. Matusitz J, Spear J. Effective doctor-patient communication: an updated examination. Soc Work Public Health 2014;29(3):252-266. [doi: 10.1080/19371918.2013.776416] [Medline: 24802220]
8. Kelley J, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
9. Hall JA, Roter DL, Rand CS. Communication of affect between patient and physician. J Health Soc Behav 1981 Mar;22(1):18-30. [Medline: 7240703]
10. Meeuwesen L, Schaap C, van der Staak C. Verbal analysis of doctor-patient communication. Soc Sci Med 1991;32(10):1143-1150. [doi: 10.1016/0277-9536(91)90091-p] [Medline: 2068597]
11. Duggan P, Parrott L. Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. Human Comm Res 2001 Apr;27(2):299-311. [doi: 10.1111/j.1468-2958.2001.tb00783.x]
12. Aruguete M, Roberts C. Participants' ratings of male physicians who vary in race and communication style. In: Psychological Reports. Los Angeles, CA: SAGE Publications; Dec 2002:793-806.
13. Scott D, Purves IN. Triadic relationship between doctor, computer and patient. Interact Comput 1996 Dec;8(4):347-363. [doi: 10.1016/s0953-5438(97)83778-2]
14. Vogel D, Meyer M, Harendza S. Verbal and non-verbal communication skills including empathy during history taking of undergraduate medical students. BMC Med Educ 2018 Jul 03;18(1):157 [FREE Full text] [doi: 10.1186/s12909-018-1260-9] [Medline: 29970069]
15. Alkureishi M, Lee W, Lyons M, Press V, Imam S, Nkansah-Amankra A, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. J Gen Intern Med 2016 May;31(5):548-560 [FREE Full text] [doi: 10.1007/s11606-015-3582-1] [Medline: 26786877]
16. Frankel R, Altschuler A, George S, Kinsman J, Jimison H, Robertson NR, et al. Effects of exam-room computing on clinician-patient communication: a longitudinal qualitative study. J Gen Intern Med 2005 Aug;20(8):677-682 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0163.x] [Medline: 16050873]
17. Chen Y, Cheng K, Tang C, Siek KA, Bardram JE. Is my doctor listening to me? impact of health IT systems on patient-provider interaction. 2013 Presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; April 27-May 2; Paris, France p. 2419-2426. [doi: 10.1145/2468356.2468791]
18. Als AB. The desk-top computer as a magic box: patterns of behaviour connected with the desk-top computer; GPs' and patients' perceptions. Fam Pract 1997 Feb 01;14(1):17-23. [doi: 10.1093/fampra/14.1.17] [Medline: 9061339]
19. McGrath J, Arar N, Pugh J. The influence of electronic medical record usage on nonverbal communication in the medical interview. Health Informatics J 2007 Jun;13(2):105-118 [FREE Full text] [doi: 10.1177/1460458207076466] [Medline: 17510223]
20. Hashim MJ. Patient-centered communication: basic skills. Am Fam Physician 2017 Jan 01;95(1):29-34 [FREE Full text] [Medline: 28075109]
21. Silverman J, Kinnersley P. Doctors' non-verbal behaviour in consultations: look at the patient before you look at the computer. Br J Gen Pract 2010 Feb 01;60(571):76-78. [doi: 10.3399/bjgp10x482293]
22. Pearce C, Arnold M, Phillips C, Trumble S, Dwan K. The patient and the computer in the primary care consultation. J Am Med Inform Assoc 2011 Mar 01;18(2):138-142 [FREE Full text] [doi: 10.1136/jamia.2010.006486] [Medline: 21262923]
23. Pearce C, Trumble S. Computers can't listen--algorithmic logic meets patient centredness. Aust Fam Physician 2006 Jun;35(6):439-442 [FREE Full text] [Medline: 16751862]
24. Noordman J, Verhaak P, van Beljouw I, van Dulmen S. Consulting room computers and their effect on general practitioner-patient communication. Fam Pract 2010 Dec 26;27(6):644-651. [doi: 10.1093/fampra/cmq058] [Medline: 20660530]
25. Crampton N, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665 [FREE Full text] [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
26. Pearce C, Trumble S, Arnold M, Dwan K, Phillips C. Computers in the new consultation: within the first minute. Fam Pract 2008 Jun;25(3):202-208. [doi: 10.1093/fampra/cmn018] [Medline: 18504254]
27. Booth N, Robinson P, Kohannejad J. Identification of high-quality consultation practice in primary care: the effects of computer use on doctor-patient rapport. Inform Prim Care 2004 May 01;12(2):75-83 [FREE Full text] [doi: 10.14236/jhi.v12i2.111] [Medline: 15319059]
28. Asan O, D Smith P, Montague E. More screen time, less face time - implications for EHR design. J Eval Clin Pract 2014 Dec 19;20(6):896-901 [FREE Full text] [doi: 10.1111/jep.12182] [Medline: 24835678]

XSL•FO
RenderX

29. Asan O, Young HN, Chewning B, Montague E. How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. Patient Educ Couns 2015 Mar;98(3):310-316 [FREE Full text] [doi: 10.1016/j.pec.2014.11.024] [Medline: 25534022]
30. Kee J, Khoo H, Lim I, Koh M. Communication skills in patient-doctor interactions: learning from patient complaints. Health Prof Educ 2018 Jun;4(2):97-106. [doi: 10.1016/j.hpe.2017.03.006]
31. Helou S, Abou-Khalil V, Yamamoto G, Kondoh E, Tamura H, Hiragi S, et al. Understanding the situated roles of electronic medical record systems to enable redesign: mixed methods study. JMIR Hum Factors 2019 Jul 09;6(3):e13812 [FREE Full text] [doi: 10.2196/13812] [Medline: 31290398]
32. Ha J, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
33. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. Ann Fam Med 2006 Mar 01;4(2):124-131 [FREE Full text] [doi: 10.1370/afm.425] [Medline: 16569715]
34. Pearce C, Dwan K, Arnold M, Phillips C, Trumble S. Doctor, patient and computer--a framework for the new consultation. Int J Med Inform 2009 Jan;78(1):32-38. [doi: 10.1016/j.ijmedinf.2008.07.002] [Medline: 18752989]
35. Chan W, Stevenson M, McGlade K. Do general practitioners change how they use the computer during consultations with a significant psychological component? Int J Med Inform 2008 Aug;77(8):534-538. [doi: 10.1016/j.ijmedinf.2007.10.005] [Medline: 18036885]
36. Lanier C, Cerutti B, Dominicé Dao M, Hudelson P, Junod Perron N. What factors influence the use of electronic health records during the first 10 minutes of the clinical encounter? Int J Gen Med 2018 Oct;11:393-398. [doi: 10.2147/ijgm.s178672]
37. Park S, Chen Y, Raj S. Beyond health literacy: supporting patient-provider communication during an emergency visit. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25- March 1; Portland, Oregon p. 2179-2192. [doi: 10.1145/2998181.2998357]
38. Antoun J, Hamadeh G, Romani M. Effect of computer use on physician-patient communication using interviews: a patient perspective. Int J Med Inform 2019 May;125:91-95. [doi: 10.1016/j.ijmedinf.2019.03.005] [Medline: 30914186]
39. Sustersic M, Gauchet A, Kernou A, Gibert C, Foote A, Vermorel C, et al. A scale assessing doctor-patient communication in a context of acute conditions based on a systematic review. PLoS One 2018 Feb 21;13(2):e0192306 [FREE Full text] [doi: 10.1371/journal.pone.0192306] [Medline: 29466407]
40. Zabar S, Ark T, Gillespie C, Hsieh A, Kalet A, Kachur E, et al. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? Acad Emerg Med 2009 Sep;16(9):915-918 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00510.x] [Medline: 19673703]
41. Street R, Liu L, Farber N, Chen Y, Calvitti A, Weibel N, et al. Keystrokes, mouse clicks, and gazing at the computer: how physician interaction with the EHR affects patient participation. J Gen Intern Med 2018 Apr;33(4):423-428 [FREE Full text] [doi: 10.1007/s11606-017-4228-2] [Medline: 29188544]
42. Roter D, Larson S. The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. Patient Educ Couns 2002 Apr;46(4):243-251. [doi: 10.1016/s0738-3991(02)00012-5] [Medline: 11932123]
43. Weibel N, Rick S, Emmenegger C, Ashfaq S, Calvitti A, Agha Z. LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office. Pers Ubiquit Comput 2014 Sep 28;19(2):317-334. [doi: 10.1007/s00779-014-0821-0]
44. Goodwin MA, Stange KC, Zyzanski SJ, Crabtree BF, Borawski EA, Flocke SA. The Hawthorne effect in direct observation research with physicians and patients. J Eval Clin Pract 2017 Dec 28;23(6):1322-1328 [FREE Full text] [doi: 10.1111/jep.12781] [Medline: 28752911]
45. Zahle J. Privacy, informed consent, and participant observation. Perspect Sci (Neth) 2017 Aug;25(4):465-487. [doi: 10.1162/posc_a_00250]
46. Zheng K, Hanauer D, Weibel N, Agha Z. Computational ethnography: automated and unobtrusive means for collecting data in situ for human-computer interaction evaluation studies. In: Cognitive Informatics for Biomedicine. Cham, Switzerland: Springer; 2015:111-140.
47. Hart Y, Czerniak E, Karnieli-Miller O, Mayo AE, Ziv A, Biegon A, et al. Automated video analysis of nonverbal communication in a medical setting. Front Psychol 2016 Aug 23;7:1130 [FREE Full text] [doi: 10.3389/fpsyg.2016.01130] [Medline: 27602002]
48. Gutstein D, Montague E, Furst J, Raicu D. Hand-eye coordination: automating the annotation of physician-patient interactions. 2019 Presented at: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30; Athens, Greece p. 657-662. [doi: 10.1109/BIBE.2019.00123]
49. Gutstein D, Montague E, Furst J, Raicu D. Optical flow, positioning, and eye coordination: automating the annotation of physician-patient interactions. 2019 Presented at: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30; Athens, Greece p. 943-947. [doi: 10.1109/bibm47256.2019.8983239]
50. Dev-Audio: innovative microphones for groups. Biamp Systems. URL: http://www.dev-audio.com/ [accessed 2021-04-28]
51. Faucett HA, Lee ML, Carter S. I should listen more: real-time sensing and feedback of non-verbal communication in video telehealth. Proc ACM Hum Comput Interact 2017 Dec 06;1(CSCW):1-19. [doi: 10.1145/3134679]

XSL•FO
RenderX

52. Comp-Ethno/gaze-dialogue-doc-pat-comp. GitHub. URL: https://git.io/JL8KJ [accessed 2021-04-28]
53. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2021 Jan 1;43(1):172-186. [doi: 10.1109/tpami.2019.2929257]
54. WebRTC real-time communication for the web. Google Developers. URL: http://www.webrtc.org/ [accessed 2021-04-28]
55. Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, et al. Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci U S A 2009 Jun 30;106(26):10587-10592 [FREE Full text] [doi: 10.1073/pnas.0903616106] [Medline: 19553212]
56. Hoey E. Lapses: How people arrive at, and deal with, discontinuities in talk. Res Lang Soc Interact 2015 Nov 18;48(4):430-453. [doi: 10.1080/08351813.2015.1090116]
57. Keisanen T, Rauniomaa M, Haddington P. Suspending action: from simultaneous to consecutive ordering of multiple courses of action. In: Multiactivity in Social Interaction: Beyond Multitasking. Amsterdam: John Benjamins Publishing Company; 2014:109-134.
58. Kendrick K, Torreira F. The timing and construction of preference: a quantitative study. Discourse Process 2015 Mar 03;52(4):255-289. [doi: 10.1080/0163853x.2014.955997]
59. Pearce C, Walker H, O'Shea C. A visual study of computers on doctors' desks. Inform Prim Care 2008;16(2):111-117 [FREE Full text] [doi: 10.14236/jhi.v16i2.682] [Medline: 18713527]
60. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009 Mar;24(3):341-348 [FREE Full text] [doi: 10.1007/s11606-008-0892-6] [Medline: 19130148]
61. Choudhury A, Crotty B, Asan O. Comparing the impact of double and single screen electronic health records on doctor-patient nonverbal communication. IISE Trans Occup Ergon Hum Factors 2020 Apr 01;8(1):42-49. [doi: 10.1080/24725838.2020.1742251]
62. Montague E, Xu J, Chen P, Asan O, Barrett B, Chewning B. Modeling eye gaze patterns in clinician-patient interaction with lag sequential analysis. Hum Factors 2011 Oct;53(5):502-516 [FREE Full text] [doi: 10.1177/0018720811405986] [Medline: 22046723]
63. Montague E, Asan O. Physician interactions with electronic health records in primary care. Health Syst (Basingstoke) 2012 Dec 01;1(2):96-103 [FREE Full text] [doi: 10.1057/hs.2012.11] [Medline: 24009982]
64. Asan O, Chiou E, Montague E. Quantitative ethnographic study of physician workflow and interactions with electronic health record systems. Int J Ind Ergon 2015 Sep 01;49:124-130 [FREE Full text] [doi: 10.1016/j.ergon.2014.04.004] [Medline: 26279597]
65. Wetterneck TB, Lapin JA, Krueger DJ, Holman GT, Beasley JW, Karsh B. Development of a primary care physician task list to evaluate clinic visit workflow. BMJ Qual Saf 2012 Jan 06;21(1):47-53 [FREE Full text] [doi: 10.1136/bmjqs-2011-000067] [Medline: 21896667]
66. Miller E, Nelson EL. Modifying the Roter Interaction Analysis System to study provider-patient communication in telemedicine: promises, pitfalls, insights, and recommendations. Telemed J E Health 2005 Feb;11(1):44-55. [doi: 10.1089/tmj.2005.11.44] [Medline: 15785220]
67. Schon D. The Reflective Practitioner: How Professionals Think In Action. New York: Basic books; 1984.
68. Sengers P, Boehner K, David S, Kaye J. Reflective design. 2005 Presented at: Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility; August 20-25; Aarhus, Denmark p. 49-58. [doi: 10.1145/1094562.1094569]
69. Liu C, Scott K, Lim R, Taylor S, Calvo R. EQClinic: a platform for learning communication skills in clinical consultations. Med Educ Online 2016;21:31801 [FREE Full text] [doi: 10.3402/meo.v21.31801] [Medline: 27476537]
70. Liu C, Lim RL, McCabe KL, Taylor S, Calvo RA. A web-based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: a randomized crossover study. J Med Internet Res 2016 Dec 12;18(9):e246. [doi: 10.2196/jmir.6299] [Medline: 27619564]
71. Kagan N, Schauble P, Resnikoff A, Danish SJ, Krathwohl DR. Interpersonal process recall. J Nerv Ment Dis 1969 Apr;148(4):365-374. [doi: 10.1097/00005053-196904000-00004] [Medline: 5768914]
72. Werner A, Schneider J. Teaching medical students interactional skills. a research-based course in the doctor-patient relationship. N Engl J Med 1974 May 30;290(22):1232-1237. [doi: 10.1056/NEJM197405302902206] [Medline: 4825853]
73. Asan O, Montague E. Technology-mediated information sharing between patients and clinicians in primary care encounters. Behav Inf Technol 2014 Apr 14;33(3):259-270 [FREE Full text] [doi: 10.1080/0144929X.2013.780636] [Medline: 26451062]
74. Chen Y, Ngo V, Harrison S, Duong V. Unpacking exam-room computing: negotiating computer-use in patient-physician interactions. 2011 Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7-12; Vancouver BC Canada p. 3343-3352. [doi: 10.1145/1978942.1979438]
75. Helou S, Abou-Khalil V, Yamamoto G, Kondoh E, Tamura H, Hiragi S, et al. Understanding the EMR-related experiences of pregnant Japanese women to redesign antenatal care EMR systems. Informatics 2019 Apr 04;6(2):15. [doi: 10.3390/informatics6020015]

XSL•FO
RenderX

76. Kanda N, Horiguchi S, Fujita Y, Xue Y, Nagamatsu K, Watanabe S. Simultaneous speech recognition and speaker diarization for monaural dialogue recordings with target-speaker acoustic models. 2019 Presented at: IEEE Automatic Speech Recognition and Understanding Workshop; December 14-18; Singapore p. 31-38. [doi: 10.1109/asru46091.2019.9004009]
77. Kodish-Wachs J, Agassi E, Kenny IIIP, Overhage JM. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. AMIA Annu Symp Proc 2018;2018:683-689 [FREE Full text] [Medline: 30815110]
78. Robinson JD. Getting down to business: talk, gaze, and body orientation during openings of doctor-patient consultations. Human Comm Res 1998 Sep;25(1):97-123. [doi: 10.1111/j.1468-2958.1998.tb00438.x]
79. Schegloff E. Body torque. Social Research 1998;65(3):535-596.

Abbreviations

D: Dialogue
SG: Screen Gaze
SG+D: Screen Gaze and Dialogue

Edited by G Eysenbach; submitted 22.10.20; peer-reviewed by L Seuren, E Bellei; comments to author 13.11.20; revised version received 07.01.21; accepted 14.04.21; published 10.05.21.

Please cite as:
Helou S, Abou-Khalil V, Iacobucci R, El Helou E, Kiyono K
Automatic Classification of Screen Gaze and Dialogue in Doctor-Patient-Computer Interactions: Computational Ethnography Algorithm Development and Validation
J Med Internet Res 2021;23(5):e25218
URL: https://www.jmir.org/2021/5/e25218
doi:10.2196/25218
PMID:33970117

©Samar Helou, Victoria Abou-Khalil, Riccardo Iacobucci, Elie El Helou, Ken Kiyono. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.05.2021. 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

Usability of Rapid Cholera Detection Device (OmniVis) for Water Quality Workers in Bangladesh: Iterative Convergent Mixed Methods Study

Theresa L Rager[1], MSc; Cristian Koepfli[1], PhD; Wasif A Khan[2], MD; Sabeena Ahmed[2], MSc; Zahid Hayat Mahmud[3], PhD; Katherine N Clayton[4], PhD

[1]Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, United States

[2]Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

[3]Laboratory Sciences and Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

[4]OmniVis Inc, South San Francisco, CA, United States

Corresponding Author:
Katherine N Clayton, PhD
OmniVis Inc
280 Utah Ave Ste 200
South San Francisco, CA,
United States
Phone: 1 415 938 4300
Email: kclayton@omnivistech.com

Abstract

Background: Cholera poses a significant global health burden. In Bangladesh, cholera is endemic and causes more than 100,000 cases each year. Established environmental reservoirs leave millions at risk of infection through the consumption of contaminated water. The Global Task Force for Cholera Control has called for increased environmental surveillance to detect contaminated water sources prior to human infection in an effort to reduce cases and deaths. The OmniVis rapid cholera detection device uses loop-mediated isothermal amplification and particle diffusometry detection methods integrated into a handheld hardware device that attaches to an iPhone 6 to identify and map contaminated water sources.

Objective: The aim of this study was to evaluate the usability of the OmniVis device with targeted end users to advance the iterative prototyping process and ultimately design a device that easily integrates into users' workflow.

Methods: Water quality workers were trained to use the device and subsequently completed an independent device trial and usability questionnaire. Pretraining and posttraining knowledge assessments were administered to ensure training quality did not confound trial and questionnaire

Results: Device trials identified common user errors and device malfunctions including incorrect test kit insertion and device powering issues. We did not observe meaningful differences in user errors or device malfunctions accumulated per participant across demographic groups. Over 25 trials, the mean time to complete a test was 47 minutes, a significant reduction compared with laboratory protocols, which take approximately 3 days. Overall, participants found the device easy to use and expressed confidence and comfort in using the device independently.

Conclusions: These results are used to advance the iterative prototyping process of the OmniVis rapid cholera detection device so it can achieve user uptake, workflow integration, and scale to ultimately impact cholera control and elimination strategies. We hope this methodology will promote robust usability evaluations of rapid pathogen detection technologies in device development.

(J Med Internet Res 2021;23(5):e22973) doi:10.2196/22973

KEYWORDS

cholera; environmental surveillance; mHealth; usability

XSL•FO
RenderX

Introduction

Cholera is a waterborne disease caused by the bacterium *Vibrio cholerae* that has led to seven major pandemics since 1817 [1]. There are currently 47 countries affected by cholera, leading to 2.9 million cases and 95,000 deaths each year [2]. Cholera is fatal in approximately 60% of untreated cases, but the fatality rate can be reduced to a less than 1% with aggressive rehydration and electrolyte replacement [3]. In total, the disease costs an estimated US $2 billion in lost productivity and health care costs annually [2]. Cholera is a disease of inequity in that it affects the poorest and most vulnerable populations within each affected country [2]. Women of childbearing age and children aged 2 to 9 years are most at risk of infection, with about half of all cases occurring in children younger than 5 years [4,5]. These inequities are attributed to access to safe water and sanitation.

Globally, 844 million people lack access to safe drinking water, and 2.4 billion are without basic sanitation facilities, leading to 2 billion people drinking water with fecal contaminants; these inequities leave billions of people vulnerable to cholera each year [2]. Community members can be asymptomatic carriers of cholera, unknowingly excreting the bacteria and potentially contaminating water supplies [6]. The persistence of *V cholerae* in community water sources leaves local populations continuously vulnerable to infection [7]. Thus, environmental water sources must be monitored for the establishment of reservoirs.

The ability of *V cholerae* to survive in viable but nonculturable states [8] poses a potential danger to public health efforts as the highly selective media used in conventional microbiology can fail to detect the organism in environmental samples [9]. Early pathogen detection is a key component to preventing infection. International efforts call for situational analysis to identify cholera hotspots, primarily through early warning surveillance and stronger laboratory capacities to reduce cholera transmission [10]. While early warning surveillance can encompass both epidemiological and environmental surveillance techniques, using the latter to monitor community water supplies and piped water infrastructure would likely detect pathogens prior to infection and subsequently prevent outbreaks. If *V cholerae* is detected, governments and aid agencies can immediately deploy rapid response teams to distribute household water treatment products and provide community education. Additionally, environmental surveillance can inform government policymaking and implementation strategies by identifying cholera hotspots [2,11]. Rapid detection and environmental surveillance of *V cholerae* could contribute to these goals.

In recent years, the use of mobile technology has dramatically increased globally. Mobile health, or mHealth, is defined by the World Health Organization as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [12]. With more than 106 mobile cellular subscriptions per 100 people worldwide, mobile technologies have surpassed 100% penetrance [13]. Mobile technology has infiltrated communities around the world, uniquely positioning mHealth to address global health concerns at a low cost. Rapid pathogen detection technologies could have a significant impact on environmental surveillance efforts [7]. Current mHealth rapid pathogen detection technologies primarily use lab-on-a-chip or microfluidic technology [14]. Because of their camera, GIS, and data storage modalities, mHealth technologies are ideally suited to detect pathogens in environmental water sources and serve as a low-cost alternative to laboratory testing in low-resource settings [14-17].

Recognizing the need for increased point-of-use laboratory technologies in the field and in low-resource settings, OmniVis Inc developed a smartphone-based rapid cholera detection device (Figure 1) [18,19].

XSL•FO
RenderX

Figure 1. OmniVis rapid cholera detection device with inserted iPhone 6 and workstation setup for laboratory personnel trials.

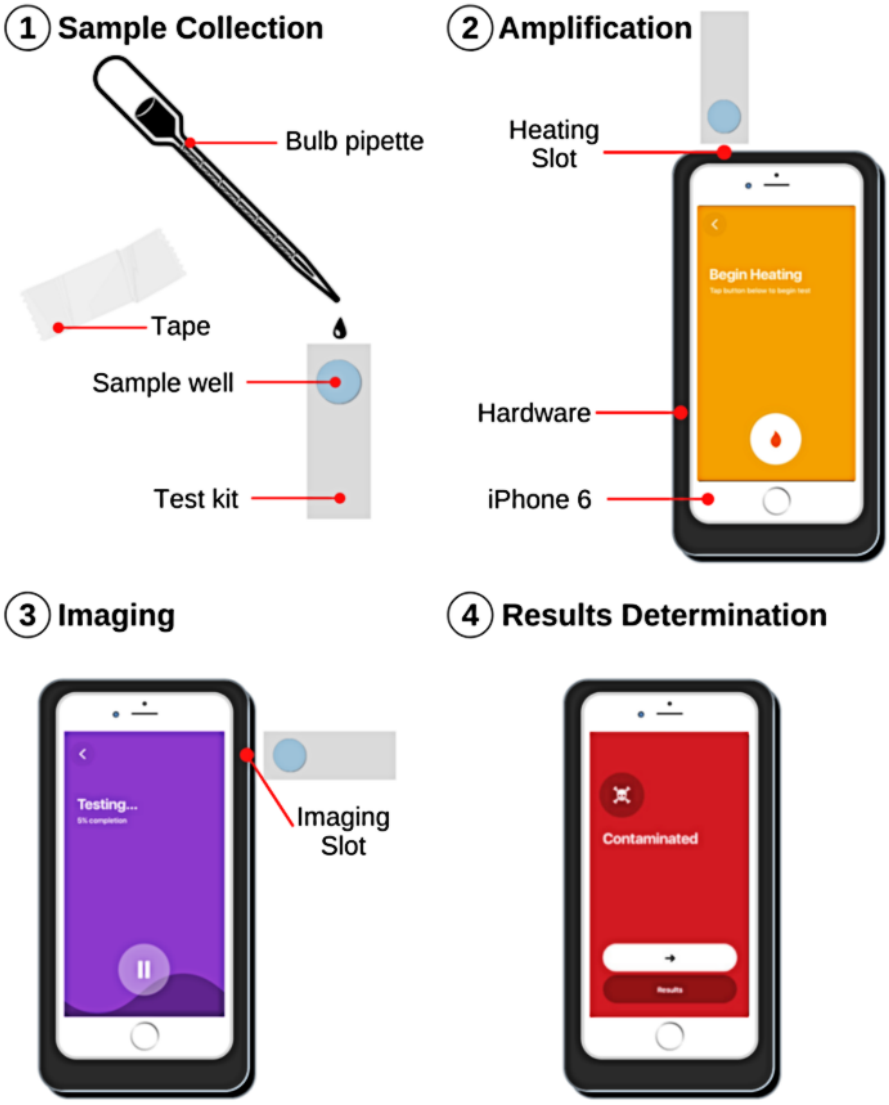


An environmental water sample (approximately 150 µL) is collected in a single-use disposable test kit and inserted into a 3D-printed hardware device with an embedded heating unit and microscope that is built to fit an iPhone 6 preloaded with the OmniVis app (Figure 2). Using particle diffusometry and loop-mediated isothermal amplification, the device detects the presence of cholera from the small water sample [18-20]. Prior to this research, the device was tested only in laboratory settings, demonstrating an estimated time-of-use of 45 minutes.

The purpose of this study was to progress the iterative prototyping process of the OmniVis rapid cholera detection device by evaluating the device's usability with targeted end users, namely water quality workers. As defined by the International Organization for Standardization, usability is the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [21]. Therefore, this study seeks to determine the effectiveness with which water quality workers can use the OmniVis device, its efficiency in detecting the presence of toxigenic cholera (*V cholerae* with the ctxA gene-encoding cholera toxin) in environmental water samples, and workers' satisfaction in using the device. While the field of mHealth and rapid diagnostic testing for infectious disease is fast-growing, few rapid pathogen detection technologies have been developed for environmental surveillance [12,22]. Moreover, many mHealth usability studies have evaluated interventions such as behavioral change, medication adherence, or clinical decision-making technologies, but there is a literature gap for usability studies evaluating rapid pathogen detection technology [23-26]. By providing input in the device development process, end users can become cocreators and feel ownership in the device design, enabling greater uptake and integration into workflows. Therefore, it is critical to conduct usability evaluations to increase device uptake and ensure the development of culturally appropriate technologies [27]. It is important to assess the usability of the OmniVis rapid cholera detection device with targeted end users in order to contribute to environmental surveillance, control, and elimination efforts.

XSL•FO
RenderX

Figure 2. OmniVis device and app workflow: (1) environmental water sample is collected into well of single-use disposable test kit and channel is sealed, (2) test kit is inserted into hardware heating unit for loop-mediated isothermal amplification, (3) after 40 minutes, sample is placed in microscope imaging slot, (4) OmniVis app displays results: contaminated/not contaminated.



Methods

Study Location

The user-centered study took place in Dhaka, Bangladesh. Dhaka has an estimated population of 14.4 million growing by 1.02% annually and a population density of 19,447 people per square kilometer [28]. Specifically, the study was conducted in partnership with the International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), which has been conducting cholera testing, research, and surveillance in Dhaka since 1960 [29].

Participants

A convenience sample of environmental microbiology laboratory personnel (n=14) and field staff (n=11) employed by icddr,b was recruited into the study during May 2019 by the director of the environmental microbiology laboratory and director of Duaripara field operations via word of mouth at staff meetings. Eligibility was based on participant age (older than 18 years) and job role (field staff or laboratory personnel). Field staff were defined as icddr,b employees responsible for collecting environmental water samples throughout Dhaka city and the surrounding areas using sterile technique and maintaining cold chain transportation from the field to icddr,b laboratories. Laboratory personnel were defined as icddr,b employees who receive environmental water samples and conduct gold standard methods including enrichment, subculturing, and DNA amplification by polymerase chain reaction of *V cholerae* and other waterborne pathogens. Exclusion criteria included minors and nonemployees of icddr,b for adequate homogeneity of job roles. As a small-scale pilot study, only 25 participants were recruited due to funding constraints for participant compensation. All participants provided informed consent before participating in the study. Device training, device trials, and administering the usability questionnaire were conducted at the Duaripara field office for field staff participants and at the environmental microbiology laboratory for laboratory personnel participants. All participants received 820 BDT (US $10) as compensation for their time. This study was approved by the University of Notre Dame institutional review board and the icddr,b research and ethics review committees.

XSL•FO
RenderX

Device Trials

Device Training

Development of a training program was not a specific objective of this research. However, participants needed to have a uniform understanding of the device prior to conducting field trials to minimize confounding variables such as familiarity with technology. Thus, prior to conducting device trials, participants were trained on how to use the device and biosafety protocols for device use. In order to identify any confounding variables resulting from poorly conducted training, the training was evaluated through a 5-question pretraining and posttraining knowledge assessment that included both true/false and open-ended questions (Multimedia Appendix 1). These assessments tested participants' comprehension of personal protective equipment (PPE) use, sample collection, sample analysis, and proper test kit disposal. Pretraining and posttraining knowledge assessments were identical in their content. Participants completed the pretraining knowledge assessment before undergoing a device training session. Following assessment completion, the study investigator demonstrated donning PPE, collecting a water sample with single-use bulb pipette, inserting and sealing the sample into a single-use test kit, powering the device, inserting the test kit into heating and imaging slots at appropriate stages, navigating the app interface, safe disposal of the test kit and sample, and doffing PPE. Spoken English to Bangla translation of this demonstration was provided by a staff member employed by icddr,b. After completion of the training, participants completed the posttraining knowledge assessment to determine if their understanding of how to use the device improved as a result of this training.

Field Trials

Upon entering the laboratory setting or field office setting (hereafter, field), participants were provided with one hardware device, one iPhone 6, one test kit and an accompanying piece of tape to seal the sample well, one single-use bulb pipette for sample collection, and one biohazard bag for sample disposal. Deionized water was used in place of an environmental water sample to reduce the risk of exposure and contamination. Participant time-of-use for each trial was recorded starting immediately prior to collecting the deionized water sample and ending with disposing the test kit into the biohazard bag. Participants performed one trial relying on instructions displayed in the iPhone app and their understanding of the training. The investigator observed participant interactions with the device and recorded written field notes and photographs. Field notes included specific errors observed during each trial. Specific errors were grouped by similarity and summed. Groups were then more broadly coded into user errors and device malfunctions depending on the cause of the error. Photographs were supplied to OmniVis Inc for internal use in developing future prototypes.

Usability Questionnaire

A usability questionnaire (separate from the pretraining and posttraining assessments) was designed to capture user-defined errors, attitudes toward the device, and potential workflow integration pathways. Specific aims included training evaluation, device usability, workflow integration, and community education. Empirically evaluated usability questionnaires including the Mobile Phone Usability Questionnaire and the Poststudy System Usability Questionnaire were analyzed for applicable content when developing the questionnaire [30,31]. Relevant questions were adapted and reworded for applicability to the OmniVis device and testing setting. Questions were carefully ordered to follow device training and use to minimize respondent confusion or recall bias. Special attention was paid to minimize participant cognitive load by ensuring that question language was clear, specific, and nonbiased; redundant questions were omitted, and display and skip logic was used. Demographic questions were added to the end of the questionnaire to capture job role, gender, language competency, education, level of experience, and years worked at icddr,b. Both the English and Bangla versions of this study's questionnaire were input into Qualtrics Surveys Offline Application and administered via a Fire 7 (Amazon) tablet, iPhone XR (Apple Inc), or iPhone 6 (Apple Inc). Participants completed the questionnaire immediately after the field trial via self-administration.

Analysis

Pretraining and posttraining knowledge assessment scores and time-of-use data were input into SPSS (version 26, IBM Corp) for statistical analysis. Field notes were input into Excel spreadsheets (Microsoft Corp) for coding and cleaning before being transferred to SPSS for statistical analysis. Questionnaire data were downloaded from Qualtrics as an SPSS data sheet for statistical analysis. The sample included 25 participants (14 laboratory personnel and 11 field staff). Nonparametric tests were used for analysis due to the limited sample size.

Device Training

Participants were given one point per correct answer on the pretraining and posttraining knowledge assessments. The change in these scores was analyzed on SPSS based on demographics of job role, gender, and language competency.

Field Trials

Field notes were coded by user errors and device malfunctions. Descriptive frequencies were obtained. Aggregate counts of errors and malfunctions were analyzed across participant job role, gender, and language competency using a chi-square test of independence (α=.05) on SPSS.

Usability Questionnaire

Descriptive frequencies for Likert response option format and dichotomous questions were obtained. Chi-square goodness-of-fit tests and Wilcoxon signed-rank tests (α=.05) were performed on SPSS to analyze the distribution of responses around a neutral point for Likert-style questions. Dichotomous questions were analyzed using a Fisher exact test (α=.05) on SPSS to determine if there was an association between demographics and user confidence and comfort with the device.

XSL•FO
RenderX

Results

Participants

The sample included 25 participants, including 56% (14/25) laboratory personnel and 44% (11/25) field staff, 52% (13/25) English speakers and 48% (12/25) non-English speakers, and 80% (20/25) men and 20% (5/25) women. Participant education levels ranged from some secondary school to master's degree (Table 1). Participant years working for icddr,b ranged from 2 months to 15 years, and their years of experience in water quality testing ranged from 3 months to 10 years. Participant demographics are further summarized in Table 1.

Table 1. Participant demographics (n=25).

Characteristic	Value, n (%)
Job role	
Field staff	11 (44)
Laboratory personnel	14 (56)
Gender	
Male	20 (80)
Female	5 (20)
Language	
English	13 (52)
Non-English	12 (48)
Education[a]	
Some secondary school	1 (4)
Secondary school	3 (12)
Some university	3 (12)
University	9 (36)
Masters	9 (36)
Years at icddr,b[b]	
<1	5 (20)
1-5	10 (40)
5-10	6 (24)
10+	4 (16)
Years of experience in water quality testing[c]	
<1	6 (25)
1-5	14 (58)
5-10	3 (13)
10+	1 (4)

[a]Questionnaire responses of some primary school, primary school, and PhD removed as no participants indicated these as the highest level of education.

[b]icddr,b: International Centre for Diarrheal Disease Research, Bangladesh.

[c]One participant declined to answer this question, hence 24 total responses.

Device Training

Pretraining and posttraining assessments were conducted with all participants, and assessment answers were analyzed. After the pretraining and posttraining assessments were scored, it appeared that participants were confused by questions 1 and 2 (Multimedia Appendix 1) because these questions were frequently incorrect on both the pretraining and posttraining knowledge assessments. Question 1 assessed recommended PPE use for device trials. A correct answer listed goggles, gloves, and gown. Question 2 assessed steps to collecting and sealing the environmental water sample into the test kit. A correct answer described use of a bulb pipette to fill the test kit's water channel. Participants expressed confusion to the investigator about differences between US and Bangladesh PPE recommendations and the wording of question 2. Thus, assessments were analyzed with and without these questions. To assess knowledge gain, changes in scores from the pretraining and posttraining assessment were analyzed across demographic features. Table 2 demonstrates that the median change in scores between the pretraining and posttraining assessment was positive (ie, scores were higher on the

XSL•FO
RenderX

postassessment) for all groups when analyzed with and without questions 1 and 2 (5 points vs 3 points). Positive changes suggest that participant knowledge pertaining to the necessary steps for sample collection, sample analysis, and biosafety protocols for the OmniVis device use were gained through the training. Similar scores across demographics suggest that participants obtained approximately the same level of overall understanding from the training regardless of job role, gender, or language competency.

Table 2. Change in assessment scores by demographics.

Change in score	Job role		Gender		Language	
	Field staff	Lab personnel	Male	Female	English	Non-English
5 points, median	2	1	2	2	2	2
3 points, median	1	1	1	1	1	1

Field Trials

User Errors and Device Malfunctions

Over the course of 25 field trials, 70 user errors and 59 device malfunctions were recorded in total (ie, not per individual user). User errors are defined as improper use of the device dependent on participant actions, and device malfunctions are defined as a failure of the device independent of participant actions. Of the 70 user errors, 21 were unique, and of the 59 device malfunctions, 16 were unique. Unique errors and malfunctions were coded into the categories displayed in Table 3. Unique errors were categorized according to the perception, cognition, action framework for better evaluation [32]. Perception errors occur when users fail to perceive a stimulus (eg, visual or tactile); cognition errors include user memory, rule-based, or knowledge-based failures; and action errors occur when users are unable to act on stimuli (eg, activating a control or applying correct force).

Table 3. Descriptive frequencies of aggregate user errors and device malfunctions.

Error class and type	Value, n (%)	Error category		
		Perception	Cognition	Action
User error (n=70)				
Incorrect tape use	5 (7)		✓	
Incorrect test kit use	12 (17)		✓	
Incorrect app use	11 (16)	✓		
Incorrect test kit insertion into imaging slot	8 (11)	✓		
Failure to perform thumb press	15 (21)			✓
Device powering	19 (27)			✓
Device malfunction (n=59)				
Unresponsive switches	8 (14)			✓
Poor test kit fit	5 (9)			✓
Poor fitting tape	18 (31)			✓
Loss of device power	27 (46)			✓
App closure	1 (2)			✓

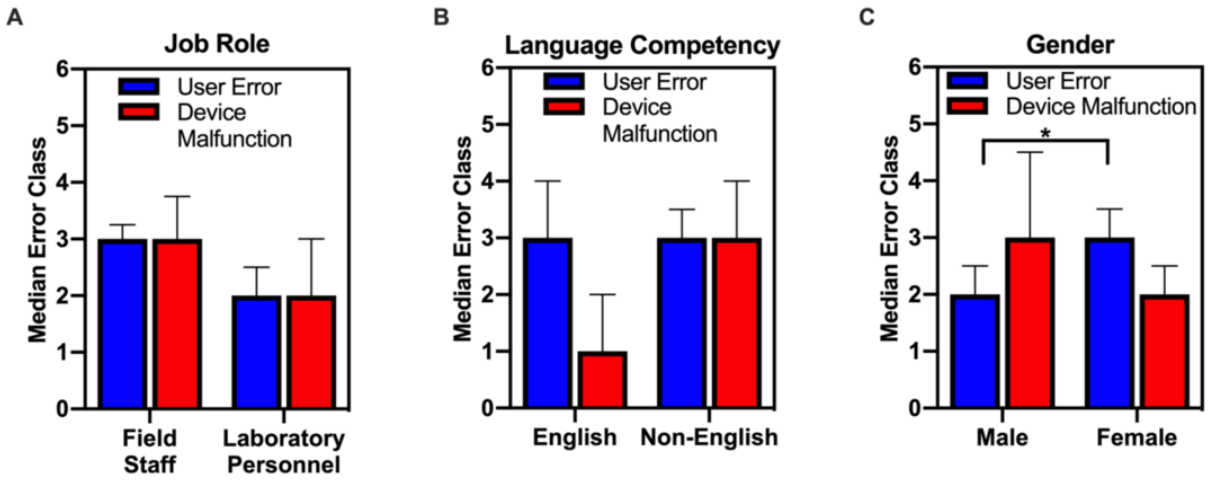
Counts of user errors and device malfunctions were compared across job role, gender, and language competency to determine if there was an association between demographics and user errors and device malfunctions. There was a significant association between language competency and error class (χ^2_1=4.2, *P*=.04), suggesting the number of user errors and device malfunctions differed between English and non-English speakers. Comparing aggregate counts across all 25 participants, non-English speakers experienced fewer user errors (non-English=30; English=40) but more device malfunctions than English speakers (non-English=36; English=23).

Further analysis was performed to determine if there was a significant difference in the median number of user errors and device malfunctions per participant according to job role, gender, and language competency (Figure 3). Nonparametric tests were used for analysis because of the limited sample size. To compare the median number of errors across demographics, Mann-Whitney *U* tests (α=.05) were performed. No statistical significance was found for median number of device malfunctions or median number of user errors between field staff and laboratory personnel (Figure 3A) or English and non-English speakers (Figure 3B). Further, no statistical significance was found for the median number of device malfunctions between men and women (Figure 3C). However, there was a statistically significant difference between the median number of user errors per participant between men and women (*U*=20.5, *P*=.04). Although this difference was

XSL•FO
RenderX

statistically significant, the median user errors for men and women differed by only one error per user. Furthermore, the Cramer *V* statistic suggests that the strength of this association is relatively weak (V=0.099). Therefore, this statistically significant yet relatively weak association may be confounded by unequal gender representation in the sample. Because the sample sizes were not equal (men=20, women=5), the median number of user errors by gender and the strength of this association should be evaluated in future studies with larger sample sizes and more equal gender representation.

Figure 3. User errors and device malfunctions compared across demographics. Statistically significant (*P=.04) difference in median user errors between men and women. Errors bars represent semiinterquartile ranges.



Time of Use

Time of use was recorded as the time required for participants to complete one trial of a mock cholera test on the OmniVis device, beginning with picking up the test kit for sample collection and ending with disposing of the test kit in the biohazard bag. Mean time of use was 46:43 minutes (min=41:58; max=53:46). This observed time of use is less than 2 minutes greater than the hypothesized time of use (45 minutes). Compared with current gold standard laboratory methodologies that require nearly 3 days to detect cholera in environmental water samples, this mean time of use demonstrates a substantial reduction in detection time [33].

Usability Questionnaire

Training Evaluation and Device Usability

Ease-of-use questions were formatted as Likert response options. English translation of these questions can be found in Multimedia Appendix 2. Question responses were coded as follows: 1=very difficult/dissimilar, 2=difficult/dissimilar, 3=neither easy/similar nor difficult/dissimilar, 4=easy/similar, 5=very easy/similar. Descriptive frequencies of these questions are shown in Table 4.

The mean and median for each question suggests that participants responded positively to the training components and device features evaluated because these descriptive frequencies are all greater than 3 (neither easy/similar nor difficult/dissimilar). Chi-square goodness of fit tests (α=.05) were performed to determine if question responses significantly differed from a uniform distribution (20% response for each answer choice). Results suggest that responses did not follow a uniform distribution and participants selected certain answer choices more than others (Table 4).

Further analysis was conducted to determine if the median response score for each question in the usability study significantly differed from 3 (neither easy/similar nor difficult/dissimilar), which represents a neutral response. A median response score greater than 3 suggests a positive response, while a response score less than 3 suggests a negative response. Wilcoxon signed-rank tests (α=.05) were performed for each Likert-style question, and results are shown in Table 4. At the 5% level of significance, the median score for each question is significantly greater than 3 (P<.001), suggesting that participants responded favorably to the training session and found the device easy to use overall.

XSL•FO
RenderX

Table 4. Training evaluation and device usability chi-square goodness of fit and 1-sample Wilcoxon signed-rank tests.

Question topic	Minimum	Median	Maximum	Chi-square value[a]	Chi-square significance	1-Sample Wilcoxon signed-rank significance[b]
Training session	3	4	5	19.2	0.001	0
Sample collection training	2	4	5	34.4	0	0
Sample seal training	2	4	5	20.0	0	0
Assembly training	2	4	5	22.0	0	0
Operate training	3	4	5	23.2	0	0
Assembly	2	4	5	20.0	0	0
Screen prompts	3	4	5	33.6	0	0
Words read	3	4	5	42.8	0	0
Color change	2	4	5	33.2	0	0
Interface similarity	2	4	5	25.6	0	0
Water collection	3	4	5	29.6	0	0
Water sample seal	3	4	5	25.6	0	0
Test kit insertion	2	4	5	15.6	0.004	0
Results read	4	4	5	42.4	0	0
Data transfer	3	4	5	42.8	0	0
Disassemble	2	4	5	29.2	0	0

[a]No cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0. Degrees of freedom for all chi-square goodness of fit tests are 4.0.

[b]Asymptotic significances are displayed for 1-sample Wilcoxon signed-rank tests. The significance level is .05.

User Confidence and Comfort

Nine dichotomous questions were designed to evaluate participant confidence and comfort using the OmniVis device. English translation of these questions can be found in Multimedia Appendix 2. Descriptive frequencies are displayed in Table 5.

Table 5. Descriptive frequencies of dichotomous questions.

Question topic	Value, n (%)	
	No	Yes
Effective test	0 (0)	25 (100)
Confidence	1 (4)	24 (96)
Device size for transport	0 (0)	25 (100)
Device size for use	0 (0)	25 (100)
Safety in the field	3 (12)	22 (88)
Safety in public	6 (24)	19 (76)
Durability	5 (20)	20 (80)
Proximity to water source	6 (24)	19 (76)
Functions	3 (12)	22 (88)

For each dichotomous question, the majority of participants responded yes, suggesting they are confident and comfortable in their ability to use the device. Participants responded most favorably to the device's size and ability to effectively test the water sample, as indicated by the 100% yes responses. Participants expressed the least confidence and comfort in using the device in public and close to a water source, as 24% of participants answered no to these questions.

Dichotomous question responses were evaluated with Fisher exact tests (α=.05) to test for significant associations according to demographics. Statistical analysis for usability questions "Can you effectively test the water sample with the device?" "Would the device's size be conducive to transporting it in the field or lab?" and "Would the device's size be conducive to using it in the field or lab?" were excluded because 100% of participants responded yes to these questions. The only statistically significant result found was between gender and

XSL•FO
RenderX

participant response to the question "Is the device durable enough to use in the field or lab?" (χ^2_1=6.25, *P*=.04). Of the female participants, 60% (3/5) responded that the device would not be durable enough to use in the field or lab while only 10% (2/20) of male participants gave the same response. Because the frequency difference between gender responses is minimal (3 female participants vs 2 male participants), these results may be confounded by a small sample size and unequal distribution of genders in the sample.

Discussion

Principal Findings

This study evaluates the usability of the OmniVis rapid cholera detection device with targeted end users in Dhaka, Bangladesh. Despite the importance of evaluating usability, many mHealth technologies fail to undergo such evaluations and subsequently fail to scale [12,27]. While many rapid pathogen detection technologies have undergone validity analyses, there is a literature gap of usability analyses for such technologies [14-18]. Many usability studies instead evaluate technologies that aid interventions such as behavioral change, treatment adherence, or clinical decision making [23-26]. Thus, this study can serve as a model of iterative convergent mixed methods research design for device development to further promote usability analysis of rapid pathogen detection technologies and produce rapid detection devices that better fit into user workflows for increased uptake. These results contribute to the iterative prototyping process of the OmniVis device, helping the device to achieve scale and contribute to cholera environmental surveillance. The device trials measured the effectiveness and efficiency of the device by defining common user errors and device malfunctions while also evaluating the length of time required to complete one test. The usability questionnaire elucidated water quality workers' satisfaction with the device with questions regarding ease of use about specific components of device training and use. The questionnaire also evaluated user confidence and comfort in using the device in various settings.

Although device training was not a specific objective of this study, it was necessary to provide uniform training to participants to ensure equal knowledge of the device prior to conducting trials, and thus, minimize confounding variables. Pretraining and posttraining knowledge assessments were administered to detect if poor understanding of training confounded device trial and usability questionnaire results. Overall, participant knowledge increased as a result of the training, as demonstrated by the positive median change in scores across demographics. The only poorly understood competency was proper PPE use. Participants frequently answered question 1 ("What 3 pieces of PPE must you put on before collecting a water sample?") incorrectly by listing different pieces of PPE other than those required by the knowledge assessment. Thus, this error suggests participants are accustomed to different PPE rather than a lack of understanding of the purpose of PPE. Nevertheless, confusion detected in assessment questions 1 and 2 indicate that the training component of this study must be reformed in order to better instruct users how to operate the device for accurate widespread use.

The objective of the device trials was to identify common user errors and device malfunctions in order to correct these errors and malfunctions in subsequent iterations of the device. Results from these device trials suggest that common user errors center around incorrect test kit use when moving from the heating to imaging phases and device powering (Figure 2, Table 3). Common device malfunctions included loss of device power and poorly fitting tape (Table 3). There was a significant association between language competency and error class in which non-English speakers experienced fewer user errors and more device malfunctions than English speakers. Because device malfunctions are defined as failures of the device independent of the participant's actions, this association likely has a temporal confounding variable. Field staff represented a large proportion of non-English speakers in this study. Since all field staff performed their trials on the same day (trials numbers 9 to 19), device malfunctions likely resulted from previous use. For example, the 8 previous trials (trials numbers 1 to 8) likely contributed to decreased battery life, which resulted in more device malfunctions in subsequent trials and contributed to the association between non-English speakers and device malfunctions. These errors elucidate tangible improvements for future device iterations, including removal of the test kit insertion stage and redesign of device powering to make the device effective and usable in field settings.

Additionally, user errors and device malfunctions were analyzed quantitatively to determine if any particular demographic experienced errors at a higher frequency than others. By considering ease of use across demographics, device development will promote equity and accessibility to all, which is critical for widespread use in low-resource settings. The difference in median user errors for men and women was statistically significant but found to have a relatively weak association (Figure 3). Although the proportion of women in the sample was quite low, such analysis provided a gender lens to ensure that the device is not designed more favorably for men or women but rather equally usable by all genders. While there were only 5 women participants in this research, our initial studies warrant further investigation to compare error frequency between men and women in future studies. There were no statistically significant differences in median user errors or device malfunctions per participant when comparing field staff and laboratory personnel or English and non-English speakers. These results suggest that device usability is not biased according to job role or language competency.

Device trials were also used to evaluate the claim that the OmniVis device can detect toxigenic *V cholerae* in an environmental water sample in 45 minutes or less. Time-of-use data suggests that the true median test time is likely greater than 45 minutes, at an average of 46 minutes and 43 seconds. However, no participant required more than 1 hour to complete the test, suggesting that the device dramatically reduces the time required to detect toxigenic *V cholerae* in an environmental water sample. By reducing the time required to determine the presence or absence of cholera in water sources, the OmniVis device can contribute to rapid detection, early warning, and

XSL•FO
RenderX

environmental surveillance of cholera in water sources—methods defined as pivotal for cholera control and elimination [10].

Usability questionnaire results clearly demonstrate that participants perceived the device use positively. Questions regarding training ease of understanding and device ease of use had median scores greater than 3, suggesting participants had positive responses (Table 4). Moreover, participants frequently responded yes to dichotomous questions, suggesting participants expressed confidence and comfort in using the device independently (Table 5). Participants responded most favorably to the device's size and its conduciveness to transport and use in the field or lab. However, safety and durability concerns were noted. Women expressed greater concern for the durability of the device in the field (P=.04). Although the difference was not significant, field staff expressed increased concern for using the device in public (P=.06). Because field staff collect environmental water samples in public more than laboratory personnel do, they could have increased concern about publicly using an expensive device, namely an iPhone, in low-resource areas or near water sources. These concerns prompt reevaluating the use of an iPhone as opposed to a lower cost mobile platform in future iterations because safety concerns could be a barrier to uptake in this population of targeted end users.

Limitations

Limitations of this study include sample size and testing conditions. The study involved only 25 participants, all of whom were employed by icddr,b. All reported a relatively high level of education and training. Participants represented targeted end users in Bangladesh but not all cholera-endemic countries, making the generalizability of the findings for other countries limited, but the intent of the study was to apply the findings to the OmniVis device rather than generalizing results to the population of water quality workers. In this sense, the findings provide insight into the usability of the device. In addition, only 5 women were included in the sample size. Efforts were made to include as many women as possible to achieve equal gender representation, but the sample still comprised an unequal gender distribution. The small sample size also limited the ability to pilot the pretraining and posttraining knowledge assessment and questionnaire (doing so would have depleted the available sample size). Although the questionnaire was based on the Poststudy System Usability Questionnaire, significant adaptations were made to apply the questionnaire to this study. Without such pilots, the reliability and validity of the knowledge assessment and questionnaire could not be evaluated. Additionally, field trials were completed with demonstration test kits that lacked the correct chemistry to complete a pathogen detection test. Because these demonstration test kits had to be reused in subsequent trials, environmental water samples were mimicked with deionized water. Thus, the conditions under which the device was evaluated do not perfectly mirror real testing conditions.

Conclusion

World public health authorities have called for the evaluation of mHealth technologies in order to improve their ability to achieve scale [21]. This usability study is the first step in gaining feedback from targeted end users on the effectiveness, efficiency, and satisfaction of the OmniVis device. By employing an iterative convergent mixed methods design, this study collects user insights into device development and incorporates targeted end user attitudes and perceptions into the development of the OmniVis device [34]. Such methodology increases the likelihood of user uptake of the rapid pathogen detection technology, further contributing to environmental surveillance efforts [27]. The study demonstrated that a portable cholera detection test can be completed in less than 1 hour (mean time of use=46:43 minutes), which is a significant time reduction from the current laboratory protocol [33]. Participants responded positively to the device training and found the device easy to use at each phase of the testing process. Moreover, participants expressed confidence and comfort in using the device independently. While these results are promising for the this prototype, concerns over durability and safety require addressing in subsequent prototypes, particularly across various demographics. Future device iterations can include removing the test kit insertion stage, redesigning device powering, and adapting to a less-expensive mobile platform. Ultimately, this study will help OmniVis further develop the rapid cholera detection device so the device can better contribute to environmental surveillance efforts for cholera control and elimination. Newer generations of the OmniVis device will undergo larger user-centered trials in additional low-resource settings such as field settings in Kenya and Haiti. These studies will provide more feedback as the device is tested in different cultural and language profiles. In addition to advancing OmniVis device development, this study also promotes a feasible methodology for evaluating the usability of rapid pathogen detection technologies. Usability evaluations will help more mHealth and point-of-use environmental surveillance devices achieve scale, integration, and interoperability to further efforts to control, eliminate, and eradicate communicable diseases.

Acknowledgments

The authors would like to thank Dr Miguel Morales for his mentorship leading up to this study and Dr Michelle Ngai for her helpful discussions. Further, we would like to show our gratitude to Mr Jordan Florian for his support and discussions in preparation for the study in Dhaka, Bangladesh, and Mr Dong Hoon Lee for device assembly and preparation. We would like to acknowledge Grand Challenges Canada (grant R-HGC-POC-1803-22784) and the Asmuth Family Foundation Fellowship for their gracious funding of this work. This research study was also funded by core donors who provide unrestricted support to icddr,b for its operations and research, including the governments of Bangladesh, Canada, Sweden, and the United Kingdom. We gratefully acknowledge our core donors for their support and commitment to icddr,b's research efforts.

XSL•FO
RenderX

Authors' Contributions

TLR had full access to all the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. TLR and KNC conceived and designed the study. TLR acquired, analyzed, and interpreted the data and drafted the manuscript. Critical revision of the manuscript was provided by CK and KNC. CK, WAK, SA, ZHM, and KNC provided administrative, technical, or material support. KNC also provided supervisory support.

Conflicts of Interest

KNC is a cofounder and full-time employee of OmniVis Inc. The other authors have no conflicts to declare.

Multimedia Appendix 1
Pretraining and posttraining knowledge assessment (English and Bangla).
[DOCX File , 15 KB - jmir_v23i5e22973_app1.docx]

Multimedia Appendix 2
Usability questionnaire (English with coding).
[DOCX File , 18 KB - jmir_v23i5e22973_app2.docx]

Multimedia Appendix 3
Usability Questionnaire (Bangla with Coding).
[DOCX File , 33 KB - jmir_v23i5e22973_app3.docx]

References

1. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of Vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio 2011;2(4):e00157 [FREE Full text] [doi: 10.1128/mBio.00157-11] [Medline: 21862630]
2. Global Task Force on Cholera Control. Ending Cholera: A Global Roadmap to 2030. Geneva: World Health Organization; 2017. URL: https://www.who.int/cholera/publications/global-roadmap.pdf?ua=1 [accessed 2020-05-22]
3. Mahalanabis D, Molla A, Sack D. Clinical management of cholera. In: Cholera, 3rd Edition. Amsterdam: Plenum Medical Book Company; 1992:253-283.
4. Ali M, Lopez AL, You YA, Kim YE, Sah B, Maskery B, et al. The global burden of cholera. Bull World Health Organ 2012 Mar 01;90(3):209-218 [FREE Full text] [doi: 10.2471/BLT.11.093427] [Medline: 22461716]
5. Glass R, Becker S, Huq M, Stoll BJ, Khan MU, Merson MH, et al. Endemic cholera in rural Bangladesh, 1966-1980. Am J Epidemiol 1982 Dec;116(6):959-970. [doi: 10.1093/oxfordjournals.aje.a113498] [Medline: 7148820]
6. Azman AS, Rudolph KE, Cummings DA, Lessler J. The incubation period of cholera: a systematic review. J Infect 2013 May;66(5):432-438 [FREE Full text] [doi: 10.1016/j.jinf.2012.11.013] [Medline: 23201968]
7. Ramamurthy T, Das B, Chakraborty S, Mukhopadhyay AK, Sack DA. Diagnostic techniques for rapid detection of Vibrio cholerae O1/O139. Vaccine 2020 Feb 29;38 Suppl 1:A73-A82 [FREE Full text] [doi: 10.1016/j.vaccine.2019.07.099] [Medline: 31427135]
8. Righetto L, Zaman RU, Mahmud ZH, Bertuzzo E, Mari L, Casagrandi R, et al. Detection of Vibrio cholerae O1 and O139 in environmental waters of rural Bangladesh: a flow-cytometry-based field trial. Epidemiol Infect 2015 Aug;143(11):2330-2342. [doi: 10.1017/S0950268814003252] [Medline: 25496520]
9. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, et al. Critical factors influencing the occurrence of Vibrio cholerae in the environment of Bangladesh. Appl Environ Microbiol 2005 Aug;71(8):4645-4654 [FREE Full text] [doi: 10.1128/AEM.71.8.4645-4654.2005] [Medline: 16085859]
10. Fourth Meeting of the Water, Sanitation and Hygiene (WASH) Working Group. Geneva: Global Task Force on Cholera Control, World Health Organization; 2019 Feb 12. URL: https://www.fondation-merieux.org/en/events/4th-meeting-of-the-global-task-force-on-cholera-control-gtfcc-wash-working-group/ [accessed 2020-05-29]
11. Marshall C, Lewis D, Whittaker M. mHealth technologies in developing countries: a feasibility assessment and a proposed framework. Health Information Systems Knowledge Hub, School of Population Health, U Queensland. 2013 Jun. URL: https://tinyurl.com/6n8dza3f [accessed 2020-05-22]
12. MHealth: New Horizons for Health through Mobile Technologies, Volume 3. Geneva: World Health Organization; 2011 Jun 07. URL: http://whqlibdoc.who.int/publications/2011/9789241564250_eng.pdf?ua=1 [accessed 2020-05-22]
13. Mobile cellular subscriptions (per 100 people). World Bank Data. 2017. URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2?end=2018&start=1960&view=chart [accessed 2020-05-22]
14. Yang K, Peretz-Soroka H, Liu Y, Lin F. Novel developments in mobile sensing based on the integration of microfluidic devices and smartphones. Lab Chip 2016;16(6):943-958. [doi: 10.1039/c5lc01524c]

XSL•FO
RenderX

15. Park TS, Li W, McCracken KE, Yoon J. Smartphone quantifies Salmonella from paper microfluidics. Lab Chip 2013;13(24):4832-4840. [doi: 10.1039/c3lc50976a]
16. Park TS, Cho S, Nahapetian TG, Yoon J. Smartphone detection of UV LED-enhanced particle immunoassay on paper microfluidics. SLAS Technol 2017 Feb;22(1):7-12 [FREE Full text] [doi: 10.1177/2211068216639566] [Medline: 26993147]
17. Zhu H, Mavandadi S, Coskun AF, Yaglidere O, Ozcan A. Optofluidic fluorescent imaging cytometry on a cell phone. Anal Chem 2011 Sep;83(17):6641-6647. [doi: 10.1021/ac201587a]
18. Colbert AJ, Co K, Lima-Cooper G, Lee DH, Clayton KN, Wereley S, et al. Towards the use of a smartphone imaging-based tool for point-of-care detection of asymptomatic low-density malaria parasitemia. Malaria J 2021 Mar 15:1-32 (forthcoming) [FREE Full text] [doi: 10.21203/rs.3.rs-335992/v]
19. Moehling TJ, Lee DH, Henderson ME, McDonald MK, Tsang PH, Kaakeh S, et al. A smartphone-based particle diffusometry platform for sub-attomolar detection of Vibrio cholerae in environmental water. Biosensors Bioelectr 2020 Nov;167:1. [doi: 10.1016/j.bios.2020.112497]
20. Clayton KN, Berglund GD, Linnes JC, Kinzer-Ursem TL, Wereley ST. DNA microviscosity characterization with particle diffusometry for downstream DNA detection applications. Anal Chem 2017 Dec 06;89(24):13334-13341. [doi: 10.1021/acs.analchem.7b03513]
21. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017 Jun;10(sup3):1327686 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
22. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
23. Samson LL, Pape-Haugaard L, Meltzer MC, Fuchs M, Schønheyder HC, Hejlesen O. Design of a tablet computer app for facilitation of a molecular blood culture test in clinical microbiology and preliminary usability evaluation. JMIR Mhealth Uhealth 2016 Mar 18;4(1):e20 [FREE Full text] [doi: 10.2196/mhealth.5041] [Medline: 26993432]
24. Compernolle S, Cardon G, van der Ploeg HP, Van Nassau F, De Bourdeaudhuij I, Jelsma JJ, et al. Engagement, acceptability, usability, and preliminary efficacy of a self-monitoring mobile health intervention to reduce sedentary behavior in belgian older adults: mixed methods study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e18653 [FREE Full text] [doi: 10.2196/18653] [Medline: 33118951]
25. Guo X, Yang Y, Takiff H, Zhu M, Ma J, Zhong T, et al. A comprehensive app that improves tuberculosis treatment management through video-observed therapy: usability study. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e17658 [FREE Full text] [doi: 10.2196/17658] [Medline: 32735222]
26. Bilal S, Nelson E, Meisner L, Alam M, Al Amin S, Ashenafi Y, et al. Evaluation of standard and mobile health-supported clinical diagnostic tools for assessing dehydration in patients with diarrhea in rural bangladesh. Am J Trop Med Hyg 2018 Jul;99(1):171-179 [FREE Full text] [doi: 10.4269/ajtmh.17-0648] [Medline: 29761756]
27. Qiang C, Yamamichi M, Hausman V, Miller R, Altman D. Mobile applications for the health sector. World Bank ICT Sector Unit. 2012. URL: http://documents1.worldbank.org/curated/en/751411468157784302/pdf/726040WP0Box370th0report00Apr020120.pdf [accessed 2020-05-22]
28. Bangladesh population 2019. World Population Review. URL: https://worldpopulationreview.com/countries/bangladesh-population/ [accessed 2020-05-22]
29. Clemens J. icddr,b Strategic Plan 2015-18—Long version. International Centre for Diarrheal Disease Research, Bangladesh. 2015. URL: https://www.icddrb.org/dmdocuments/icddrb%20SP%202015-18%20LONG%20VERSION%20FOR%20WEB.pdf [accessed 2020-05-22]
30. Ryu Y. Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods [Dissertation]. Charlottesville: Virginia Tech University Libraries; Jul 09, 2005.
31. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
32. Redmill F. Human Factors in Safety Critical Systems. Oxford: Butterworth-Heinemann Books; 1997.
33. Alam M, Hasan NA, Sadique A, Bhuiyan NA, Ahmed KU, Nusrin S, et al. Seasonal cholera caused by Vibrio cholerae serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. Appl. Environ Microbiol 2006 Jun 02;72(6):4096-4104. [doi: 10.1128/aem.00066-06]
34. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656. [doi: 10.2196/11656] [Medline: 31025951]

Abbreviations

icddr,b: International Centre for Diarrheal Disease Research, Bangladesh
mHealth: mobile health
PPE: personal protective equipment

XSL•FO
RenderX

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by I Gabashvili, I Schiering, T Alessa; comments to author 02.12.20; revised version received 22.01.21; accepted 04.04.21; published 12.05.21.

Please cite as:
Rager TL, Koepfli C, Khan WA, Ahmed S, Mahmud ZH, Clayton KN
Usability of Rapid Cholera Detection Device (OmniVis) for Water Quality Workers in Bangladesh: Iterative Convergent Mixed Methods Study
J Med Internet Res 2021;23(5):e22973
URL: https://www.jmir.org/2021/5/e22973
doi:10.2196/22973
PMID:33978590

©Theresa L Rager, Cristian Koepfli, Wasif A Khan, Sabeena Ahmed, Zahid Hayat Mahmud, Katherine N Clayton. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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

Patient Recruitment Into a Multicenter Clinical Cohort Linking Electronic Health Records From 5 Health Systems: Cross-sectional Analysis

Wendy L Bennett[1], MPH, MD; Carolyn T Bramante[2], MPH, MD; Scott D Rothenberger[3], PhD; Jennifer L Kraschnewski[4], MPH, MD; Sharon J Herring[5], MPH, MD; Michelle R Lent[6], PhD; Jeanne M Clark[1], MPH, MD; Molly B Conroy[7], MPH, MD; Harold Lehmann[1], MD, PhD; Nickie Cappella[3], BS; Megan Gauvey-Kern[1], BA, MS; Jody McCullough[4], BA; Kathleen M McTigue[3], MS, MPH, MD

[1]Johns Hopkins University School of Medicine, Baltimore, MD, United States
[2]University of Minnesota School of Medicine, Minneapolis, MN, United States
[3]University of Pittsburgh, Pittsburgh, PA, United States
[4]Penn State College of Medicine, Hershey, PA, United States
[5]Lewis Katz School of Medicine, Philadelphia, PA, United States
[6]Geisinger Health, Danville, PA, United States
[7]University of Utah School of Medicine, Salt Lake City, UT, United States

Corresponding Author:
Wendy L Bennett, MPH, MD
Johns Hopkins University School of Medicine
2024 E Monument St
Rm 2-616
Baltimore, MD, 21205
United States
Phone: 1 4107465146
Email: wendy.bennett@jhmi.edu

Abstract

Background: There is growing interest in identifying and recruiting research participants from health systems using electronic health records (EHRs). However, few studies have described the practical aspects of the recruitment process or compared electronic recruitment methods to in-person recruitment, particularly across health systems.

Objective: The objective of this study was to describe the steps and efficiency of the recruitment process and participant characteristics by recruitment strategy.

Methods: EHR-based eligibility criteria included being an adult patient engaged in outpatient primary or bariatric surgery care at one of 5 health systems in the PaTH Clinical Research Network and having ≥2 weight measurements and 1 height measurement recorded in their EHR within the last 5 years. Recruitment strategies varied by site and included one or more of the following methods: (1) in-person recruitment by study staff from clinical sites, (2) US postal mail recruitment letters, (3) secure email, and (4) direct EHR recruitment through secure patient web portals. We used descriptive statistics to evaluate participant characteristics and proportion of patients recruited (ie, efficiency) by modality.

Results: The total number of eligible patients from the 5 health systems was 5,051,187. Of these, 40,048 (0.8%) were invited to enter an EHR-based cohort study and 1085 were enrolled. Recruitment efficiency was highest for in-person recruitment (33.5%), followed by electronic messaging (2.9%), including email (2.9%) and EHR patient portal messages (2.9%). Overall, 779 (65.7%) patients were enrolled through electronic messaging, which also showed greater rates of recruitment of Black patients compared with the other strategies.

Conclusions: We recruited a total of 1085 patients from primary care and bariatric surgery settings using 4 recruitment strategies. The recruitment efficiency was 2.9% for email and EHR patient portals, with the majority of participants recruited electronically. This study can inform the design of future research studies using EHR-based recruitment.

(J Med Internet Res 2021;23(5):e24003) doi:10.2196/24003

XSL•FO
RenderX

KEYWORDS

electronic health record; recruitment methods; cohort study design; recruitment; health system; bariatric; surgery; clinical research network; primary care; cohort; enrollment; research; process; efficiency; eligibility

Introduction

Background

Recruitment of patients for research using electronic health records (EHRs) has potential to enhance the applicability and efficiency of patient-centered research [1,2]. Despite logistical and ethical challenges with recruiting for research using clinical data sources (eg, EHRs) or clinical communication methods (eg, EHR patient portals) [3], researchers are developing strategies that are responsive to both health system and patient stakeholders [4]. In addition, there is interest in designing research studies that include "real-world evidence" (ie, health care information representative of patients, populations, and health care delivery systems from actual clinical settings [5,6]).

In 2014, the National Patient-Centered Research Network (PCORnet) was launched with funding from the Patient-Centered Outcomes Research Institute (PCORI). PCORnet's infrastructure can support EHR-based recruitment and study implementation [7]. The PaTH Clinical Research Network (CRN) originally brought together 4 academic medical centers in the mid-Atlantic region of the United States to build the infrastructure to share EHR data across health systems so that patient-centered clinical questions could be answered in real-world settings [8]. A fifth academic health center (Geisinger Health System) was added to the network in 2015.

Using EHR data from these 5 health systems, standardized to PCORnet specifications [9], we developed a cohort of primary care and bariatric surgery patients called the "Healthy Lifestyles, Body Weight, and Health Care" cohort. The overall aim of the cohort was to complete online surveys about quality of life, healthy lifestyles, and weight management, which could be linked to selected EHR data. Another explicit goal of the CRN was to utilize several different recruitment methods in order to compare sample demographics and survey response rates by recruitment strategy. In this paper, we describe the enrollment process and then compare 4 different recruitment strategies: (1) in-person recruitment by study staff from clinical sites, (2) US postal mail recruitment letters, (3) secure email, and (4) direct EHR recruitment through secure patient web portals. The objective of this paper was to describe the steps and efficiency of the recruitment process and participant characteristics by strategy.

Methods

Institutional review board approval was obtained from PaTH's single institutional review board at The Johns Hopkins University School of Medicine, Baltimore, Maryland. We benefitted from the input of patient stakeholders in the design of the recruitment methods and online survey. Participants who enrolled in this study were not offered any compensation.

EHR-Based Participant Eligibility Criteria

We recruited adult patients from primary care and bariatric surgery clinics to complete surveys and form an EHR cohort from 5 health care systems included in the PaTH CRN—Geisinger Health System, Johns Hopkins Health System, Penn State Milton S. Hershey Medical Center, Temple Health System, and the University of Pittsburgh Medical Center. We identified participants using EHR-based eligibility criteria (ie, the "computable phenotype"). Eligible patients were aged ≥18 years and had a minimum of 2 weight measurements and 1 height measurement recorded between January 1, 2011, and May 31, 2015. Patients in all BMI categories were eligible. Participants were excluded if they were deceased or non–English proficient, as assessed at the time of consent, because the consent form and survey were only available in the English language.

Participant Recruitment and Setting

Each health system used one or more of the following 4 EHR-based research invitation strategies: (1) in-person recruitment by study staff from partnering outpatient primary care and bariatric surgery clinics; (2) US postal mail recruitment invitation letters using the mailing address listed in the EHR, with an online registration link contained in the invitation; (3) email to potentially eligible participants using the email address stored in the EHR, with an embedded registration link in the email; and (4) secure EHR recruitment messages through the patient web portal with an embedded registration link.

Each health system tailored its own strategy to recruit participants from the large pool of potentially eligible patients identified using the computable phenotype. At each site, local norms for integrating research with clinical practice were followed, leading to differences in the number of patients approached by each strategy per site. Importantly, each site's study team partnered with specific clinical practices so that recruitment letters were jointly sent from clinical representatives involved in the patients' care and members of the research team and focused on patients from certain practices. Notably, 2 health systems used direct EHR recruitment through secure patient web portals (sites C and D), while 1 site (site B) was unable to use email because email address was not a data field within the EHR. The content of the recruitment letters and in-person scripts had similar language across all health systems and recruitment strategies.

Each health system had a slightly different recruitment window, but the total recruitment window for all sites was from April 2015 to November 2016. We stopped the recruitment when we met or exceeded the a priori goal of recruiting 1000 participants.

Consent and Enrollment Process

Regardless of the recruitment strategy or method used to complete the baseline survey, all participants completed a web-based consent form, including an initial "consent quiz" designed to ensure that patients understood that they were

XSL•FO
RenderX

agreeing to participate in a research study. Participants consented to complete online surveys and for the research study to access their EHR. Enrollment was determined by consenting and having any data entered into the baseline survey. Participants who were recruited in person used an electronic tablet to access the consent form and complete the baseline survey. For participants recruited through the EHR patient portal, the survey was embedded in the EHR, which required that it had to be completed and "submitted" in order to be accessible and counted as enrolled. We created an enrollment flag in the EHR to identify participants who were enrolled in the cohort study.

Data Collection

Participants completed a baseline survey about sociodemographic background, weight management practices, weight-related interactions with the health system, diet, physical activity, and quality of life using standard survey measures [10-12].

We extracted EHR data (eg, laboratory values, blood pressure, anthropometric measurements, diagnosis, and procedure codes) from each site's PCORNet Common Data Model for all enrolled participants [7,13]. The EHR and survey data were linked for analysis.

Statistical Methods

For this paper, we were primarily interested in describing the patterns of enrollment by recruitment modality, study site, and participant characteristics.

Anthropomorphic data were cleaned to remove unlikely values (eg, BMI ≤15 kg/m^2 or ≥90 kg/m^2; height ≤4 ft or ≥7 ft; weight ≤50 lbs or ≥700 lbs) prior to analyses. We identified the BMI in the EHR data closest to the enrollment date, within a maximum of 3 years, and used a 5-year window to determine comorbidity diagnoses based on International Statistical Classification of Diseases and Related Health Problems codes [14].

We used descriptive statistics (means and standard deviations for continuous measures, frequencies and proportions for nominal measures) to examine differences between the large pool of participants deemed eligible using the computable phenotype and those who enrolled in the study. We defined the "efficiency" of each recruitment strategy as the proportion of the total numbers of participants who enrolled divided by the total number of participants who were approached by the strategy (eg, total enrolled via email divided by total who were sent the email). We performed ANOVA *F* tests and chi-square tests (or Fisher exact tests, where appropriate) to determine if participant characteristics varied by recruitment modality. Because the 5 health systems preferentially designed and utilized different recruitment modalities, we were not able to assess predictors of enrollment by modality using regression models because of complete collinearity by study site. Two-sided $P \leq .05$ was considered statistically significant. Analyses were performed using SAS statistical software (version 9.4; SAS Institute).

Results

Figure 1 shows the enrollment steps and efficiency of enrollment into the cohort by recruitment modality. We identified 5,051,187 eligible patients using the computable phenotype applied to the EHRs from 5 health systems. Based on each health systems' own method of partnering with clinics and recruiting participants, a total of 40,048 patients (33,839 patients from primary care clinics and 6209 patients from bariatric surgery clinics) were then sent recruitment messages or approached in person. Across all sites, the patient recruitment strategies were deployed as follows: 442 (1.1%) patients in person; 12,710 (31.7%) patients by postal mail; 25,224 (63.0%) patients by email; and 1672 (4.2%) patients by EHR patient portals. A total of 1185 participants were enrolled in the cohort, with 907 (76.5%) from primary care practices and 278 (23.5%) from bariatric surgery practices. The efficiency of enrollment by recruitment strategy was by far the highest for in-person recruitment (148/442, 33.5%), followed by email (730/25,224, 2.9%) and EHR patient portal (49/1672, 2.9%), with postal mail being least efficient (258/12,710, 2.0%). Overall, 65.7% (779/1185) were enrolled through electronic messaging (email or EHR portal).

XSL•FO
RenderX

Figure 1. Recruitment flow of participants enrolled into the cohort from 5 health systems using 4 recruitment strategies. EHR: electronic health record.

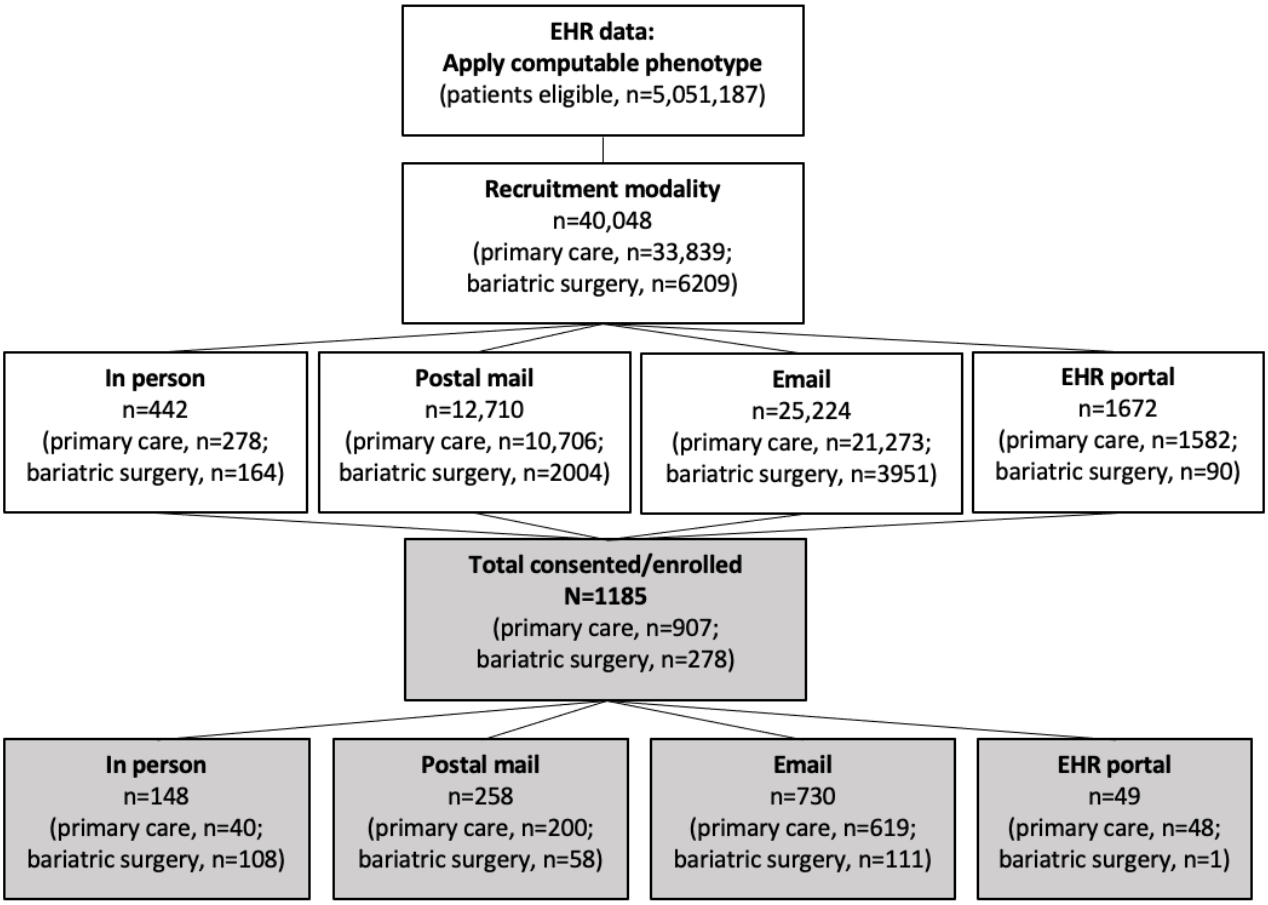


Table 1 compares the demographics and medical conditions between the 1185 participants enrolled in the cohort from primary care and bariatric surgery clinics and the 5,051,187 total patients deemed potentially eligible based on the application of the computable phenotype to their EHRs. Compared with all eligible participants, enrollees were older (aged 58.1 years in primary care clinics and 56.5 years in bariatric surgery clinics versus 43.6 years among all eligible). The bariatric surgery group had the highest proportion of women (207/278, 74.5%) compared with primary care (595/907, 65.6%) and all eligible patients (2,778,178/5,051,187, 55.0%). The proportion of Black individuals among the enrolled patient groups was smaller than that among all patients deemed eligible (81/907, 8.9% enrolled in primary care and 27/278, 9.7% in bariatric surgery versus 558,125/5,051,187, 11.1% from all patients deemed eligible). Overall, very few Hispanic participants were enrolled (14/907, 1.5% in primary care and 1/278, 0.4% in bariatric surgery) compared with the number of Hispanic individuals among the health systems' eligible participants (127,045/5,051,187, 2.7%). Information regarding level of education was not available for nonenrolled patients; in those enrolled from either primary care or bariatric surgery clinics, more than 85% of participants had at least some college education. The mean BMI was highest for the patients from bariatric surgery clinics (34.3 kg/m^2), followed by the patients from primary care clinics (30.2 kg/m^2) and the total population of eligible patients (28.3 kg/m^2). A high proportion of enrollees had hypertension (507/1185, 42.8%) and diabetes (232/1185, 19.6%), but these conditions were even more common among the large sample of eligible patients who were not enrolled (689,925/5,051,187, 52.1% with hypertension and 407,036/5,051,187, 30.7% with diabetes).

XSL•FO
RenderX

Table 1. Description of participants enrolled in the study cohort from primary care and bariatric surgery clinics versus all eligible participants.[a]

Participant characteristics	Participants enrolled in the study cohort (N=1185)		
	Primary care clinics (n=907)	Bariatric surgery clinics (n=278)	All eligible participants (n=5,051,187)
Age (years), mean (SD)	58.1 (16.0)	56.5 (15.7)	43.6 (23.1)
Sex (female), n (%)	595 (65.6)	207 (74.5)	2,778,178 (55.0)
Race[b], n (%)			
American Indian/Alaska Native	3 (0.3)	1 (0.4)	10,075 (0.2)
Asian	1 (0.1)	4 (1.4)	79,876 (1.6)
Black/African American	81 (8.9)	27 (9.7)	558,125 (11.1)
White	788 (86.9)	234 (84.2)	3,812,584 (75.5)
Unknown	13 (1.4)	10 (3.6)	462,955 (9.2)
Ethnicity[b], n (%)			
Hispanic	14 (1.5)	1 (0.4)	127,045 (2.7)
Not Hispanic	865 (95.4)	267 (96.0)	4,169,412 (89.8)
Unknown	23 (2.2)	6 (2.2)	345,957 (6.8)
Education level, n (%)			
Less than high school	6 (0.6)	3 (1.1)	N/A[c]
High school graduate or GED[d]	93 (10.3)	30 (10.8)	N/A
Some college or 2-year degree	225 (24.8)	69 (24.8)	N/A
College graduate	214 (23.6)	64 (23.0)	N/A
More than college degree	361 (39.8)	104 (37.4)	N/A
No response	8 (0.9)	8 (2.9)	N/A
BMI, mean (SD)	30.2 (8.95)	34.3 (9.02)	28.3 (7.50)
BMI category, n (%)			
<18.5	7 (0.8)	0 (0.0)	227,768 (6.6)
18.5-25	256 (28.5)	39 (14.0)	1,047,409 (30.3)
25-30	260 (29.0)	62 (22.3)	998,279 (28.9)
30-35	158 (17.6)	60 (21.6)	630,232 (18.2)
35-40	96 (10.7)	41 (14.7)	310,980 (9.0)
>40	120 (13.4)	76 (27.3)	241,391 (7.0)
Comorbid health conditions, n (%)			
Heart failure	32 (3.5)	10 (3.6)	164,420 (12.4)
Hypertension	370 (40.8)	137 (49.3)	689,925 (52.1)
Diabetes	168 (18.5)	64 (23.0)	407,036 (30.7)
Health care visit in last 6 months, n (%)	707 (77.9)	213 (76.6)	887,235 (67.0)

[a]No statistical testing was performed because of overlap between the eligible group and those enrolled in the cohort.

[b]For participants enrolled in the cohort, race and ethnicity information was self-reported; for nonenrolled participants, these data were obtained from the electronic health record.

[c]N/A: not available.

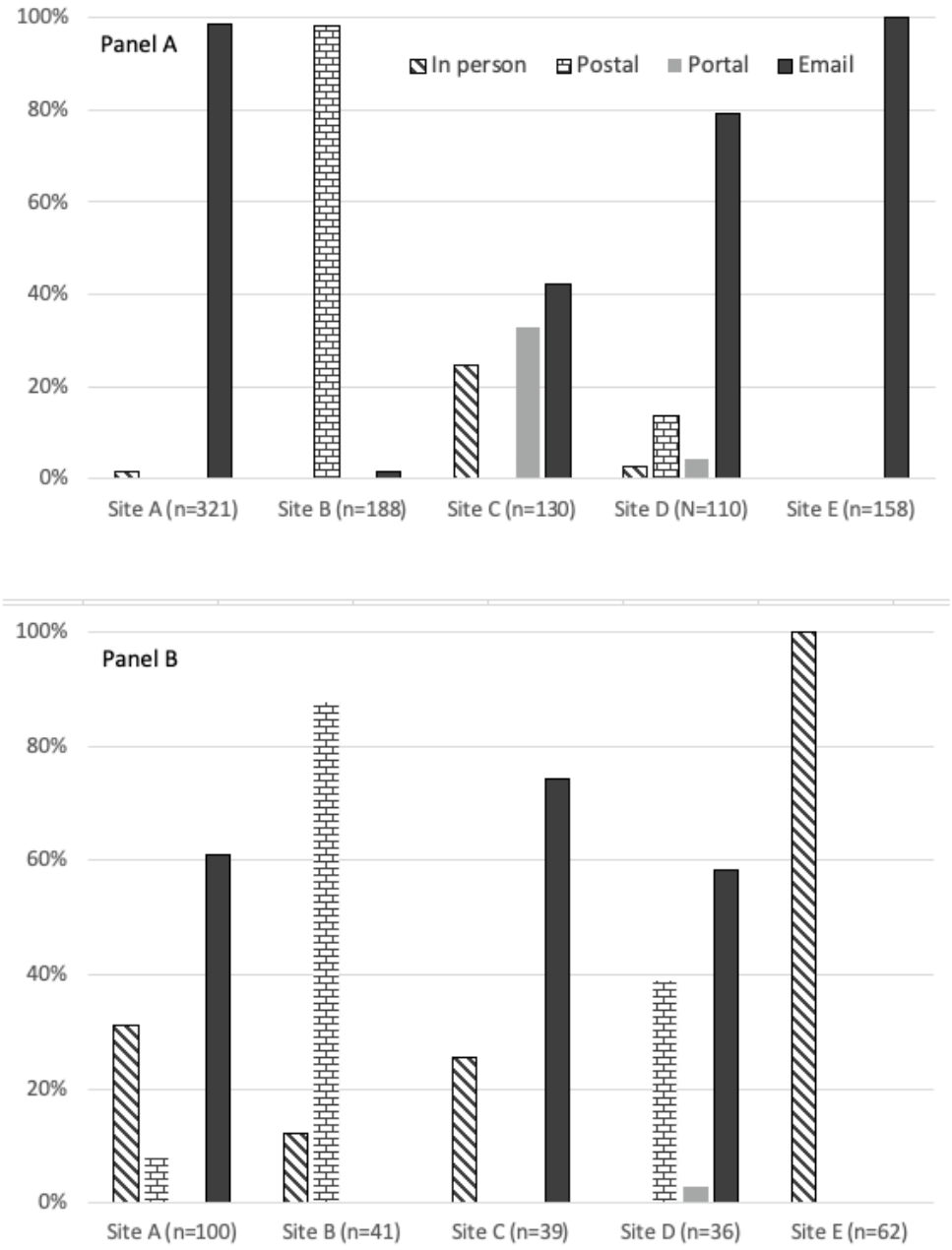
[d]GED: general education degree.

Figure 2 demonstrates the distribution of participants enrolled by strategy and recruitment site from the 5 health systems. Because of differences in research policies, staffing, and capabilities, each health system differed in the number of recruitment strategies used in the primary care and bariatric surgery settings. Table 2 shows the characteristics of enrolled cohort participants by recruitment strategy. The proportion of Black patients who were recruited using EHR patient portals

XSL•FO
RenderX

(8/49, 16.3%) and email (80/730, 11.0%) was higher than that with in-person (7/148, 4.7%) and postal (13/258, 5.0%) strategies. We found no statistically significant differences in participants' level of education or comorbid health conditions between the 4 different recruitment strategies.

Figure 2. Distribution of participants by recruitment site (sites A to E) and recruitment strategy. (A) Primary care clinic participants (n=907). (B) Bariatric surgery clinic participants (n=278).



XSL•FO
RenderX

Table 2. Description of enrolled participants (N=1185) by recruitment strategy.

	Recruitment strategy				
Participant characteristics	In-person (n=148)	Postal (n=258)	Portal (n=49)	Email (n=730)	*P* value
Age (years), mean (SD)	58.6 (16.0)	59.0 (14.6)	60.0 (12.9)	57.1 (15.9)	.22[a]
Sex (female), n (%)	114 (77.0)	165 (64.0)	33 (67.4)	540 (74.0)	.007[b]
Race, n (%)					<.001[c]
American Indian/Alaska Native	1 (0.7)	1 (0.4)	0 (0)	2 (0.3)	
Asian	2 (1.4)	0 (0)	3 (6.1)	7 (1.0)	
Black	7 (4.7)	13 (5.0)	8 (16.3)	80 (11.0)	
White	136 (91.9)	233 (90.3)	35 (71.4)	618 (84.7)	
Other	0 (0.0)	6 (2.3)	0 (0)	14 (1.9)	
Ethnicity, n (%)					.69[c]
Hispanic	1 (0.7)	5 (1.9)	0 (0)	9 (1.2)	
Not Hispanic	142 (95.9)	249 (96.5)	49 (100)	692 (94.8)	
Education level, n (%)					.03[c]
Less than high school	1 (0.7)	5 (6.1)	1 (2.0)	0 (0)	
High school graduate or GED[d]	12 (8.1)	25 (9.7)	2 (4.1)	78 (10.7)	
Some college or 2-year degree	37 (25.0)	67 (26.0)	11 (22.4)	181 (24.8)	
College graduate	39 (26.4)	61 (23.6)	7 (14.3)	172 (23.6)	
More than college degree	58 (39.2)	101 (39.1)	28 (57.1)	290 (39.7)	
BMI, mean (SD)	32.3 (8.2)	30.7 (7.9)	31.2 (9.0)	31.4 (9.0)	.20[a]
Comorbid health conditions, n (%)					
Heart failure	7 (4.7)	6 (2.3)	1 (2.0)	28 (3.8)	.58[c]
Hypertension	62 (41.9)	110 (42.6)	21 (42.9)	315 (43.2)	.99[b]
Diabetes	32 (21.6)	50 (19.4)	12 (24.5)	138 (18.9)	.50[b]
Visit to health care provider in last 6 months, n (%)	101 (68.2)	190 (73.6)	38 (77.6)	581 (79.6)	.01[b]
Primary care clinic participant, n (%)	40 (27.0)	200 (77.5)	48 (98.0)	619 (84.8)	<.001[b]
Bariatric surgery clinic participant, n (%)	108 (73.0)	58 (22.5)	1 (2.0)	111 (15.2)	
Site, n (%)					<.001[c]
A	36 (24.3)	8 (3.1)	0 (0)	377 (51.9)	
B	5 (3.4)	221 (85.7)	0 (0)	3 (0.4)	
C	42 (28.4)	0 (0)	43 (87.8)	84 (11.6)	
D	3 (2.0)	29 (11.2)	6 (12.2)	108 (14.9)	
E	62 (41.9)	0 (0)	0 (0)	158 (21.7)	

[a]ANOVA *F* test.

[b]Chi-square test.

[c]Fisher exact test.

[d]GED: general education degree.

Discussion

This study reports the experience of the PaTH CRN's recruitment of patients from primary care and bariatric surgery clinics from 5 health systems into a study cohort. Even though the in-person recruitment had the greatest efficiency (33.5%), this strategy required research staff to identify when patients had upcoming appointments and to be available on site.

XSL•FO
RenderX

Ultimately, only 442 patients were approached in person and 148 enrolled. The majority of participants were recruited using email (730/1185, 61.6%). Although email recruitment required less staff research time, this strategy was limited by having patients' email addresses recorded and available in the EHRs. The secure patient portal was a strategy employed by 2 health systems and had a recruitment efficiency comparable to email (2.9% of those who received a message enrolled in the study). The electronic recruitment strategies may be particularly desirable during times when research staff cannot be present in the clinic, such as under COVID-19–related restrictions.

Traditionally, researchers have relied on postal mailings, newspaper or radio advertising, or random-digit dialing methods for survey-based research as well as for recruitment into clinical trials [15]. EHRs provide an avenue for screening and then directly targeting recruitment at potentially eligible patients for research studies [4,15]. Recruitment of patients from EHRs has the potential to be more inclusive by including patients who might be sicker and those who are undergoing care in real-world health care systems [5,16]. However, no method of recruitment is without the potential for selection biases. With EHR-based recruitment, there are concerns about external validity or generalizability to nonpatient or health care populations, particularly for prevention-oriented studies [17]. In addition, patients with reduced access to care or those who are not regularly followed in the health care setting may be less likely to be approached in person (ie, at the time of a clinic visit) or to receive recruitment messages when they do not have an email address or access the patient portal. While 90% of US adults now use the internet, older adults and Black citizens still show lower rates of internet access [18], but technology usage rates are increasing [19]. A recent study by Walker and colleagues [18] showed that fewer Black patients and patients over the age of 70 had access to an inpatient portal. Notably, in our study, which used a combination of recruitment strategies, approximately 9% of enrolled participants identified as Black compared with 11% among patients identified as potentially eligible. In fact, researchers could leverage the EHR to specifically target patient populations by demographics or medical diagnosis (eg, by race, ethnicity, age, or rare health conditions [20,21]).

Despite the increasing need for effective and less staff-intensive methods of recruitment for clinical trials and surveys, few studies have compared recruitment strategies [22,23]. One study reported and compared response rates from various recruitment methods (postal survey, postal invitation to complete an internet survey, and postal invitation for a telephone survey) for an environmental survey but did not include EHR-based recruitment methods [23]. The study showed the highest response rate (30%) for the telephone survey [23]. The EHR provides a new way to identify potential participants for research studies by applying electronic eligibility criteria to sometimes very large pools of potentially eligible patients [24].

A systematic review by Lai and Afseth [25] assessed the effectiveness and efficiency of EHR-based recruitment methods. They identified 13 articles, of which 11 reported recruitment efficiency and most used alerts sent directly to physicians or staff to notify of participant eligibility [25]. EHRs have multiple functionalities to support recruitment [26]. In our study, we utilized the EHR for several purposes: to generate lists of potentially eligible patients using a computable phenotype, to obtain postal mailing addresses and email addresses, and to send recruitment invitations using the secure patient portal at 2 sites. Some academic research centers have designed patient portal recruitment services to enable, but also limit, recruitment using the portal for certain approved studies [1,21,27,28]. In a 2019 single-institution study that included 13 separate EHR-based recruitment strategies using the patient portal recruitment service, the average response rate for patient portal messages was 2.9%, which was the same as our enrollment rate for both email and patient portal recruitment [21]. Interestingly, we offered no compensation to patients to enroll in our study, yet the studies reviewed by Miller et al [21] did offer compensation.

Although the computable phenotype enabled our teams to identify a very large number of patients who were potentially eligible, each health system designed its own outreach methods to patients by targeting specific clinical sites, thereby greatly reducing the potential number of patients that could have been contacted about this research opportunity. Importantly, institutional review boards at each site prohibited the "cold calling" of patients (ie, directly contacting potential research participants based on prior knowledge of the patients' health information in the absence of a treatment or clinical relationship [29]). Therefore, each site partnered with primary care and bariatric surgery providers, who approached the patients about the study first, either with their signature on the recruitment invitation or through an in-person introduction. However, a recent landscape analysis by McHugh and colleagues [29] highlighted that this universally applied "ban on cold calling" could impose a gatekeeping function and potentially reduce patient autonomy, decreasing access to research and risking the introduction of selection bias into research studies. They suggested alternative approaches to ensuring patient privacy [29], with the goal of broadening access to health research participation [15].

We identified several limitations of this study. First, this was a 5-center, multisite study and the deployed recruitment strategies were dependent on the norms for research recruitment at each site, with the implication that not all samples were directly comparable by site or by strategy. For example, 3 sites did not allow research recruitment using the EHR patient portal and 1 site did not have access to email addresses in the EHRs. Because of the differences in clinical populations' demographics between sites, there was high correlation between the site's choice of recruitment methods and the patients who were recruited, limiting our ability to draw conclusions about whether specific recruitment modalities are more (or less) effective for specific racial/ethnic groups or different age categories. Second, the overall response rate was low; however, it is very comparable to response rates from other studies that relied on EHR-based or email-based recruitment. We were able to assess for selection biases by comparing those patients who were deemed eligible with those who enrolled in the study. We showed lower uptake among Latinx patients, but this was in part because the consent and survey were limited to participants able to read in English.

XSL•FO
RenderX

Third, although this study was low burden for participants, as they consented to having the research team review medical records and complete a 20-minute online survey, we did not offer any compensation for their time, which could have limited enrollment. Therefore, we are not able to draw conclusions about how effective our recruitment strategies would be for more intensive studies, for studies not using online data collection, or for studies offering participant incentives. Fourth, we were unable to estimate cost or cost-effectiveness of the recruitment strategies or describe in detail the staffing time or corresponding resources involved for each health system, such as the effort for the health informatics team to support EHR-based research or the time for research staff to conduct in-person clinic-based recruitment. Fifth, recruitment for this study occurred approximately 4 years ago (2015-2016) and it is possible that response rates and methods for electronic recruitment have improved over time. Sixth, the 5 health systems were all based in the mid-Atlantic region of the United States, which could limit the generalizability of these findings to other regions of the country and outside the United States.

The major implication of our study is to inform the selection of recruitment strategies for the design of future cohort studies, utilizing the capabilities of the modern-day EHR system. We anticipate that other researchers could find this information useful in the design of their recruitment strategies and to estimate the expected yields from "low touch" strategies that require less personnel contact with potential participants. As health systems and institutional review boards become more comfortable and familiar with EHR-based recruitment, it will be possible to achieve greater consistency between the recruitment processes and even sample selection to reduce biases across systems. Ultimately, to facilitate EHR-based research across multiple settings (eg, inpatient and ambulatory), large health systems will need to invest in infrastructure to support and link smaller clinical centers and subsidize their use of a common EHR. An example of a successful academic and community-based network is the OCHIN network, which provides a research infrastructure using EHR data from a national network of smaller community health centers [30].

In conclusion, we recruited a total of 1085 patients from primary care and bariatric surgery clinics to complete a survey and participate in an EHR-based cohort study using 4 recruitment methods. The greatest recruitment yield was achieved using the email-based method, but the greatest efficiency resulted from in-person recruitment. Implementation of low-resource recruitment approaches has important implications for future patient-centered studies in health system settings.

Acknowledgments

Funding for the study was provided by the Patient-Centered Outcomes Research Institute.

Conflicts of Interest

None declared.

References

1. Gleason KT, Ford DE, Gumas D, Woods B, Appel L, Murray P, et al. Development and preliminary evaluation of a patient portal messaging for research recruitment service. J Clin Transl Sci 2018 Feb;2(1):53-56 [FREE Full text] [doi: 10.1017/cts.2018.10] [Medline: 31660218]
2. Heerman WJ, Jackson N, Roumie CL, Harris PA, Rosenbloom ST, Pulley J, et al. Recruitment methods for survey research: Findings from the Mid-South Clinical Data Research Network. Contemp Clin Trials 2017 Nov;62:50-55. [doi: 10.1016/j.cct.2017.08.006] [Medline: 28823925]
3. van Staa T, Goldacre B, Buchan I, Smeeth L. Big health data: the need to earn public trust. BMJ 2016 Jul 14;354:i3636. [doi: 10.1136/bmj.i3636] [Medline: 27418128]
4. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017 Jan;106(1):1-9 [FREE Full text] [doi: 10.1007/s00392-016-1025-6] [Medline: 27557678]
5. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
6. Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA 2018 Sep 04;320(9):867-868. [doi: 10.1001/jama.2018.10136] [Medline: 30105359]
7. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
8. Amin W, Tsui FR, Borromeo C, Chuang CH, Espino JU, Ford D, PaTH network team. PaTH: towards a learning health system in the Mid-Atlantic region. J Am Med Inform Assoc 2014;21(4):633-636 [FREE Full text] [doi: 10.1136/amiajnl-2014-002759] [Medline: 24821745]
9. Forrest CB, McTigue KM, Hernandez AF, Cohen LW, Cruz H, Haynes K, et al. PCORnet® 2020: current state, accomplishments, and future directions. J Clin Epidemiol 2021 Jan;129:60-67 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.09.036] [Medline: 33002635]

XSL•FO
RenderX

10. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System. URL: https://www.cdc.gov/brfss/index.html [accessed 2021-05-10]
11. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. Value Health 2019 May;22(5):537-544 [FREE Full text] [doi: 10.1016/j.jval.2019.02.004] [Medline: 31104731]
12. Paxton AE, Strycker LA, Toobert DJ, Ammerman AS, Glasgow RE. Starting the conversation performance of a brief dietary assessment and intervention tool for health professionals. Am J Prev Med 2011 Jan;40(1):67-71. [doi: 10.1016/j.amepre.2010.10.009] [Medline: 21146770]
13. Patient Centered Outcomes Research Institute. PCORnet Common Data Model (CDM) Specification, Version 5.1. URL: https://pcornet.org/wp-content/uploads/2019/09/PCORnet-Common-Data-Model-v51-2019_09_12.pdf [accessed 2021-05-10]
14. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Geneva, Switzerland: WHO; 1992.
15. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev 2018 Feb 22;2:MR000013 [FREE Full text] [doi: 10.1002/14651858.MR000013.pub6] [Medline: 29468635]
16. Patel A, Billot L. Reality and Truth: Balancing the Hope and the Hype of Real-World Evidence. Circulation 2017 Jul 18;136(3):260-262. [doi: 10.1161/CIRCULATIONAHA.117.029233] [Medline: 28716830]
17. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]
18. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the Digital Divide: Age and Race Disparities in Use of an Inpatient Portal. Telemed J E Health 2020 May;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
19. Pew Research Center. Internet/Broadband Fact Sheet. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-05-10]
20. Whicher D, Philbin S, Aronson N. An overview of the impact of rare disease characteristics on research methodology. Orphanet J Rare Dis 2018 Jan 19;13(1):14 [FREE Full text] [doi: 10.1186/s13023-017-0755-5] [Medline: 29351763]
21. Miller H, Gleason K, Juraschek S, Plante T, Lewis-Land C, Woods B, et al. Electronic medical record-based cohort selection and direct-to-patient, targeted recruitment: early efficacy and lessons learned. J Am Med Inform Assoc 2019 Nov 01;26(11):1209-1217 [FREE Full text] [doi: 10.1093/jamia/ocz168] [Medline: 31553434]
22. Maghera A, Kahlke P, Lau A, Zeng Y, Hoskins C, Corbett T, et al. You are how you recruit: a cohort and randomized controlled trial of recruitment strategies. BMC Med Res Methodol 2014 Sep 27;14:111 [FREE Full text] [doi: 10.1186/1471-2288-14-111] [Medline: 25260762]
23. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012 Aug 31;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]
24. Aroda VR, Sheehan PR, Vickery EM, Staten MA, LeBlanc ES, Phillips LS, D2d Research Group. Establishing an electronic health record-supported approach for outreach to and recruitment of persons at high risk of type 2 diabetes in clinical trials: The vitamin D and type 2 diabetes (D2d) study experience. Clin Trials 2019 Jun;16(3):306-315 [FREE Full text] [doi: 10.1177/1740774519839062] [Medline: 31007049]
25. Lai YS, Afseth JD. A review of the impact of utilising electronic medical records for clinical research recruitment. Clin Trials 2019 Apr;16(2):194-203. [doi: 10.1177/1740774519829709] [Medline: 30764659]
26. Schreiweis B, Trinczek B, Köpcke F, Leusch T, Majeed RW, Wenk J, et al. Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. Int J Med Inform 2014 Nov;83(11):860-868. [doi: 10.1016/j.ijmedinf.2014.08.005] [Medline: 25189709]
27. Plante TB, Gleason KT, Miller HN, Charleston J, McArthur K, Himmelfarb CD, STURDY Collaborative Research Group. Recruitment of trial participants through electronic medical record patient portal messaging: A pilot study. Clin Trials 2020 Feb;17(1):30-38 [FREE Full text] [doi: 10.1177/1740774519873657] [Medline: 31581836]
28. Baucom RB, Ousley J, Poulose BK, Rosenbloom ST, Jackson GP. Case report: patient portal versus telephone recruitment for a surgical research study. Appl Clin Inform 2014;5(4):1005-1014 [FREE Full text] [doi: 10.4338/ACI-2014-07-CR-0059] [Medline: 25589913]
29. McHugh KR, Swamy GK, Hernandez AF. Engaging patients throughout the health system: A landscape analysis of cold-call policies and recommendations for future policy change. J Clin Transl Sci 2018 Dec;2(6):384-392 [FREE Full text] [doi: 10.1017/cts.2019.1] [Medline: 31402985]
30. Cottrell EK, Hendricks M, Dambrun K, Cowburn S, Pantell M, Gold R, et al. Comparison of Community-Level and Patient-Level Social Risk Data in a Network of Community Health Centers. JAMA Netw Open 2020 Oct 01;3(10):e2016852 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16852] [Medline: 33119102]

XSL•FO
RenderX

Abbreviations

CRN: Clinical Research Network
EHR: electronic health record
PCORI: Patient-Centered Outcomes Research Institute
PCORnet: National Patient-Centered Research Network

Edited by G Eysenbach; submitted 02.09.20; peer-reviewed by S Bonn, B Schreiweis, H Yemm, K Hek; comments to author 11.12.20; revised version received 04.02.21; accepted 04.04.21; published 27.05.21.

Please cite as:
Bennett WL, Bramante CT, Rothenberger SD, Kraschnewski JL, Herring SJ, Lent MR, Clark JM, Conroy MB, Lehmann H, Cappella N, Gauvey-Kern M, McCullough J, McTigue KM
Patient Recruitment Into a Multicenter Clinical Cohort Linking Electronic Health Records From 5 Health Systems: Cross-sectional Analysis
J Med Internet Res 2021;23(5):e24003
URL: https://www.jmir.org/2021/5/e24003
doi:10.2196/24003
PMID:34042604

©Wendy L Bennett, Carolyn T Bramante, Scott D Rothenberger, Jennifer L Kraschnewski, Sharon J Herring, Michelle R Lent, Jeanne M Clark, Molly B Conroy, Harold Lehmann, Nickie Cappella, Megan Gauvey-Kern, Jody McCullough, Kathleen M McTigue. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

Emotional Reactions to Cybersecurity Breach Situations: Scenario-Based Survey Study

Sanja Budimir[1,2], PhD; Johnny R J Fontaine[1], PhD; Nicole M A Huijts[3], PhD; Antal Haans[3], PhD; George Loukas[4], PhD; Etienne B Roesch[5], PhD

[1]Department of Work, Organization and Society, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

[2]Department for Psychotherapy and Biopsychosocial Health, Faculty of Health and Medicine, Danube University Krems, Krems an der Donau, Austria

[3]Eindhoven University of Technology, Eindhoven, Netherlands

[4]University of Greenwich, London, United Kingdom

[5]Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

Corresponding Author:
Sanja Budimir, PhD
Department of Work, Organization and Society
Faculty of Psychology and Educational Sciences
Ghent University
Henri Dunantlaan 2
Ghent
Belgium
Phone: 32 09 264 64 56
Email: sanja.budimir@ugent.be

Abstract

Background: With the ever-expanding interconnectedness of the internet and especially with the recent development of the Internet of Things, people are increasingly at risk for cybersecurity breaches that can have far-reaching consequences for their personal and professional lives, with psychological and mental health ramifications.

Objective: We aimed to identify the dimensional structure of emotion processes triggered by one of the most emblematic scenarios of cybersecurity breach, the hacking of one's smart security camera, and explore which personality characteristics systematically relate to these emotion dimensions.

Methods: A total of 902 participants from the United Kingdom and the Netherlands reported their emotion processes triggered by a cybersecurity breach scenario. Moreover, they reported on their Big Five personality traits, as well as on key indicators for resilient, overcontrolling (internalizing problems), and undercontrolling (aggression) personality types.

Results: Principal component analyses revealed a clear 3-dimensional structure of emotion processes: emotional intensity, proactive versus fight/flight reactions, and affective versus cognitive/motivational reactions. Regression analyses revealed that more internalizing problems (β=.33, *P*<.001), resilience (β=.22, *P*<.001), and agreeableness (β=.12, *P*<.001) and less emotional stability (β=–.25, *P*<.001) have significant predictive value for higher emotional intensity. More internalizing problems (β=.26, *P*<.001), aggression (β=.25, *P*<.001), and extraversion (β=.07, *P*=.01) and less resilience (β=–.19, *P*<.001), agreeableness (β=–.34, *P*<.001), consciousness (β=–.19, *P*<.001), and openness (β=–.22, *P*<.001) have significant predictive value for comparatively more fight/flight than proactive reactions. Less internalizing problems (β=–.32, *P*<.001) and more emotional stability (β=.14, *P*<.001) and aggression (β=.13, *P*<.001) have significant predictive value for a comparatively higher salience for cognitive/motivational than affective reactions.

Conclusions: To adequately describe the emotion processes triggered by a cybersecurity breach, two more dimensions are needed over and above the general negative affectivity dimension. This multidimensional structure is further supported by the differential relationships of the emotion dimensions with personality characteristics. The discovered emotion structure could be used for consistent predictions about who is at risk to develop long-term mental well-being issues due to a cybersecurity breach experience.

(J Med Internet Res 2021;23(5):e24879) doi:10.2196/24879

XSL•FO
RenderX

KEYWORDS

cybersecurity breach victims; emotions; personality; mental health; Internet of Things

Introduction

Background

The increasing number of Internet of Things devices (IoT) and their diverse application in private and work lives offer unlimited possibilities for a connected life. However, this has also extended the scope of security breaches and cybercriminal behavior [1]. As cyberattacks became more and more focused on specific companies and individual users [2,3], they increasingly create technological, economic, social, and psychological challenges. Because of the deep penetration of IoT in personal lives, cybersecurity breaches on such devices can have far-reaching personal consequences. Work and livelihood can be disturbed, personal and social spheres can be altered, and these changes can sometimes be irrevocable. The most direct psychological effects of such events, which are intrinsically relevant to the personal goals of the user, are the emotional responses they elicit [4,5]. A leading security company reported that negative emotions including anger, annoyance, frustration, upset, and a feeling of being cheated are the common reactions to being a victim of cybercrime [6]. These emotional experiences could develop into long-term, far-reaching psychological turmoil [7-10]. Despite their central role in psychological well-being, very little is known about emotional reactions in the context of cybersecurity breaches. In this study, we (1) explore victims' emotion processes by employing a scenario study with a cybersecurity breach on a smart security camera, which is one of the most telling examples of invasion of privacy by unauthorized entrance in the private sphere [11,12], (2) explore which personality characteristics predict interindividual differences in emotional reactions to this cybersecurity breach; and (3) designed the explorative research in such a way to generate replicable findings.

Emotion Processes

In emotion research, participants are often asked to report on their own emotions by evaluating emotion and affect terms (eg, the frequently used Positive and Negative Affect Schedule [PANAS] [13]). While this type of research can generate very interesting findings, it does not allow researchers to unearth the emotion processes these affect terms refer to. To get a comprehensive view of the emotion processes that can be elicited by cybersecurity breaches, emotions are currently studied on the basis of the componential emotion approach [14]. This approach has emerged as an overarching conceptual framework within the scientific field of emotion research. According to this approach, emotions are conceptualized as processes that are elicited by goal-relevant events and consist of an interplay between 5 major components: appraisals, action tendencies, bodily responses, expressions, and subjective feelings [5]. Each component has a function. Appraisals are the evaluation of the eliciting event against one's goals, needs, and values. Action tendencies refer to the preparation and direction of adaptive action. Bodily responses refer to physiological changes that prepare the body for actual action. Expressions are the facial, vocal, and gestural reactions through which the ongoing emotion process is communicated. Through subjective feelings, the individual becomes aware of the ongoing emotion process. These feelings are often communicated with the use of emotion and affect terms. Moreover, emotion processes are evolutionary-shaped processes that have evolved from reflex-like reactions to dynamic processes open to regulation [5,14]. All aspects of the emotion process can be regulated, from the impulsive reactions to the cognitive evaluations. Having flexible emotion processes allows us to better adapt to our environment [15].

This componential emotion approach is especially promising for studying emotional experiences, as it has been demonstrated across cultural and linguistic groups that the 5 components, as well as regulation are encoded in daily language. First in 3 samples from the United Kingdom, Switzerland, and Belgium [16] and later in 31 additional samples from 24 countries and representing 20 languages (such as Chinese and Japanese) [14], it was demonstrated that 142 emotion features representing the 5 emotion components and regulation systematically constitute the meaning of 24 frequently used emotion terms [14]. The componential emotion approach forms not only a comprehensive theoretical framework but also represents how people naturally think and talk about their emotions.

Thus, to fully understand emotion dynamics, it is important to go beyond feeling and emotion terms and study all emotion components and regulation processes. In this study, the dimensions that structure the emotion processes elicited by a cybersecurity breach of a smart security camera are exploratively identified by taking all emotion components as well as regulation into account.

Person Characteristics and Emotional Reactions

To better understand the emotion dimensions involved in this scenario, we evaluate whether and how characteristics of personality are related to the reported emotional experience. To this end, we have worked with 2 broad personality models that have been shown in the past to relate to emotional functioning: the Big Five personality model [17] and the resilient/overcontrolled/undercontrolled personality type model [18,19].

Big Five Personality Model

In the first model, personality is described by the Big Five broad personality traits: extraversion, emotional stability, conscientiousness, agreeableness, and openness [17]. These traits have been shown to relate to the duration of emotional states and frequency of specific emotional experiences [20]. A very common finding is that extraversion is positively associated with positive affect and emotional stability negatively with negative affect [21]. Additionally, associations of personality traits with emotion regulation were demonstrated in several studies [22-24]. For instance, extraversion, conscientiousness, and openness were predictive for problem solving and cognitive

XSL•FO
RenderX

restructuring, while agreeableness was predictive for social support and cognitive restructuring [22-24].

Resilient/Overcontrolled/Undercontrolled Personality Type Model

The second personality model classifies people into 3 broad personality types [18,19]. Resilient people are characterized by a tendency to effectively adapt to changes and have the ability to recover well from stress and negative emotional arousal. Overcontrolled people are introverted and emotionally sensitive but also dependable. They are more likely to experience sadness and fear and are at risk of developing internalizing complaints such as depression and anxiety. Undercontrolled people are low on agreeableness and conscientiousness and high on aggressiveness and delinquency. They are more likely to experience anger and are at risk to develop externalizing problems.

Gender and Age

Next to these personality predictors, we have also looked at the relationships with gender and age. Overall gender differences in emotional reactions have been observed, with females having more intense emotional reactions compared with males [25-28]. Regarding age, a general decrease of negative affective experiences [29] and increase of healthier emotion regulation strategies [30-32] have been observed throughout the life span.

Replicability

In light of the replicability crisis in psychology [33] and because of the explorative nature of this research with the innovation to root the study of emotional experiences in the componential emotion approach, the study was executed in two samples from different countries (United Kingdom and the Netherlands) speaking different languages (English and Dutch). Moreover, participants in each country received at random one of two versions of the cybersecurity breach scenario. In one version, the smart security camera showed obvious signs of a cybersecurity breach (nonambiguous condition), and in the other version, it showed unclear signs (the ambiguous condition) that could also potentially be caused by other factors (eg, a bug in the software). By adding the latter scenario, the ecological validity of the research was increased, as in daily life it is also often unclear whether or not a dysfunction of internet-connected devices is due to a cybersecurity breach.

Methods

Sample

A total of 1045 participants were recruited through Qualtrics panel, 524 participants from the United Kingdom and 521 participants from the Netherlands. Before the data analyses, participants showing signs of not properly answering questions were removed. One of the strongest indicators that the validity of responses is at stake is nondifferentiation of the responses [34]. All participants who gave the same response on at least 75% of the GRID items and on 70% of the International Personality Item Pool 50 (IPIP-50) questionnaire items deviated from most participants in scale use and were removed (n=143 deleted cases). This left 902 participants for the analyses. Sample characteristics are presented in Table 1.

Table 1. Study sample characteristics (n=902).

Characteristics	Country of residence, n (%)		Total, n (%)
	United Kingdom (n=435)	Netherlands (n=467)	
Gender			
Female	221 (50.8)	231 (49.5)	452 (50.1)
Male	214 (49.2)	236 (50.5)	450 (49.9)
Condition			
Ambiguous	217 (49.9)	241 (51.6)	458 (50.8)
Nonambiguous	218 (50.1)	226 (48.4)	444 (49.2)

Procedure

The Qualtrics project team organized and coordinated data collection. They recruited samples from both countries based on their Qualtrics panel of participants. Quotas for samples were predefined and balanced by country of residence, gender, and scenario with age range limited from 18 to 65 years. An online questionnaire, located on the Qualtrics survey platform, was presented to participants remotely by sending them a survey link. Each participant electronically signed an online informed consent form prior to completing the questionnaire. Participants had the opportunity to complete the questionnaire within a 1-week period. The average duration of questionnaire completion was 15 minutes. Each participant was presented with an introduction explaining what IoT devices are and specifically what a smart security camera is. This was followed by the presentation of one of the two scenarios (ambiguous or nonambiguous, see complete instructions in Multimedia Appendix 1). Each participant thus evaluated only randomly assigned one scenario.

Measures

Emotion Assessment

Participants were asked to imagine they experienced one out of two cybersecurity breach scenarios. Scenario 1, which represented the ambiguous condition, was formulated as follows: “Imagine that you bought a smart security camera for your home. After some time, you notice that the shutter on your smart security camera starts opening and closing without your instruction, several times for a few minutes, then it stops for a

XSL•FO
RenderX

minute and starts again opening and closing several times and then it stops." In the nonambiguous condition (scenario 2), the formulation was "Imagine that you bought a smart security camera for your home. After some time, you notice that the shutter on your smart security camera opens without your instruction and the camera rotates toward you and then starts following your movement."

Participants were subsequently asked to report the emotional reactions they would have in the situation presented using the Cybersecurity GRID questionnaire (Multimedia Appendix 2). This is an adjusted version of the GRID instrument that was used to study the meaning of emotion words across cultural and linguistic groups [14] and is based on the componential emotion approach [5] including the assessment of the 5 emotion components and emotion regulation. In order to determine and operationalize relevant features of the emotion processes in the specific context of cybersecurity breaches, we executed a preliminary qualitative survey. In this survey, 130 participants reported on their real or expected emotional reactions in cybersecurity breach situations (either from first-hand experience or based on a third-party experience). Participants' reports included a brief description of the cybersecurity breach situation and the emotional reactions they had or would have had in that situation (referring to each of the 5 emotion components and regulation). The new Cybersecurity GRID questionnaire was based on those emotion features that were reported by at least 15% of the participants. The Cybersecurity GRID contains 76 items (19 appraisals, 16 action tendencies, 11 bodily reactions, 8 expressions, 14 subjective feelings, and 8 emotion regulation strategies). Each emotion feature was evaluated on the 7-point Likert scale commonly used in survey research ranging from 1 (strongly disagree) to 7 (strongly agree) [35].

IPIP-50

IPIP-50 [17] is a validated instrument that measures the Big Five personality factors. Participants rated how accurately each statement described them on a 5-point Likert scale ranging from 1 (very inaccurate) to 5 (very accurate). Person mean-centered scores were calculated for IPIP items and reversed according to instructions. Each factor showed good to very good internal consistency (Cronbach alpha): extraversion: α=.85, agreeableness: α=.83, conscientiousness: α=.79, emotional stability: α=.83, and for openness: α=.72.

Depression, Anxiety, and Stress Scale–21 Item

The Depression, Anxiety, and Stress Scale–21 Item (DASS-21) assesses internalizing problems, which is a key feature of the overcontrolled personality type. It is a shortened 21-item version of the Depression, Anxiety, and Stress Scale [36]. Items are rated on a 4-point scale ranging from 0 (does not apply at all) to 3 (applies very much). The total DASS-21 sum scores showed high internal consistency, α=.96.

Short-Form Buss-Perry Aggression Questionnaire

The Short-Form Buss-Perry Aggression Questionnaire [37-39] assesses aggression, which a key feature of the undercontrolled personality type. It is a short version of the Buss-Perry Aggression Questionnaire [40] and consists of 21 items rated on a 5-point scale ranging from 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me). The scale showed a high internal consistency, α=.92.

Ego Resilience Scale

The Ego Resilience Scale [41,42] is a short, revised version of the Ego-Resiliency Scale [43], measuring self-reported resilience on 10 items on a 4-point scale ranging from 1 (does not apply at all) to 4 (applies very strongly). The Cronbach alpha of the total score was .78.

Ethical Approval

Ethical approval was obtained from the ethical committee of Ghent University, Faculty of Psychology and Educational Sciences in 2017 (number 2016/67).

Results

Internal Structure of Emotional Reactions

Principal component analyses were applied to identify the major dimensions of variability among 76 emotion features. To avoid confusion between emotion components from a substantive point of view and principal components obtained from analysis, the latter will be referred to as dimensions in the remainder of the text.

To identify the number of dimensions, 3 criteria were used: (1) the scree plot based on the Eigenvalues (Multimedia Appendix 1), (2) interpretability, and (3) replicability for each language, scenario, and gender (Multimedia Appendix 1). The theoretically best interpretable rotation was selected. A highly stable and well-interpretable 3-dimensional structure was identified that accounted for 48% of the total variance (see Table 2 for the highest loading features on each dimension and Multimedia Appendix 1 for the full loading matrix).

On the first dimension, accounting for 31% of variance, all emotion features have a positive loading, with the subjective experiences loading highest (eg, I would feel panic, I would feel upset). The higher participants score on this dimension, the more intense negative emotion processes are elicited by the scenario. Therefore, this dimension is named emotional intensity.

The second dimension, accounting for 12% of variance, is a bipolar dimension. One pole is defined by proactive action tendencies to deal with the cybersecurity breach (eg, I would want to regain control over the device/account, I would want to find a solution and fix the problem). The other pole is defined by fight/flight action tendencies (eg, I would want to take revenge, I would want to isolate myself physically) and features from other components that indicate distress (eg, I would have pain in the chest). Therefore, this dimension is labeled proactive versus fight/flight.

XSL•FO
RenderX

Table 2. Results from principal component analysis of the Cybersecurity GRID questionnaire (n=920)[a].

GRID items	Dimension loading		
	1	2	3
Dimension 1: Emotional intensity			
SF5. I would feel panic.	.73	.20	–.30
SF4. I would feel afraid.	.72	.09	–.26
SF7. I would feel worried.	.70	–.15	–.21
SF6. I would feel upset.	.70	.04	–.27
SF14. I would feel uncomfortable.	.67	–.14	–.12
SF11. I would feel angry.	.67	.01	–.15
Dimension 2: Proactive versus fight/flight			
AT14. I would want to destroy whatever was close.	.32	.65	.10
AT15. I would want to take revenge.	.36	.63	.16
BR4. I would have pain in the chest.	.49	.61	–.22
AT1. I would want to stop what was happening.	.43	–.62	.16
AT9. I would want to find a solution and fix the problem.	.34	–.64	.12
AT2. I would want to regain control over the device/account.	.44	–.68	.14
Dimension 3: Affective versus cognitive/motivational			
A19. I would think “It is not safe that this device is connected to the internet.”	.58	–.07	.43
A7. I would think “My trust is betrayed.”	.58	.13	.42
A12. I would think “It is happening because someone is trying to hack and take control over my count.”	.56	–.10	.41
E8. I would be walking around nervously.	.60	.32	–.34
E7. I would be restless (touching face, hair, biting nails, nervously kicking with legs).	.58	.34	–.35
ER3. I would try to calm myself down (eg, by breathing deeply).	.59	–.09	–.37

[a]The 6 highest loadings are presented, and the full loading matrix can be found in Multimedia Appendix 1.

The third dimension, accounting for 5% of the variance, is also bipolar. All appraisal and action tendency features (eg, I would think “It is not safe that this device is connected to the Internet”) have a nonnegative loading, while all subjective experience, bodily reaction, expression, and regulation features (eg, I would try to calm myself down) have a nonpositive loading on this dimension. This dimension is labeled affective versus cognitive/motivational.

Predictors of Emotional Reactions

The scores on each of the 3 identified emotion dimensions were regressed on the personality characteristics. As the Big Five indicators and the resilience, overcontrolled, and undercontrolled indicators show both theoretical and empirical overlap (and the differences and similarities between personality models do not form the focus of this research), their predictive value was investigated separately. Hierarchical linear regression analyses were performed. In the baseline model (model 1), the predictors are country of residence, scenario, gender, and age (with United Kingdom, ambiguous situation, and women being the reference categories). In the second model the personality characteristics were added as predictors: the Big Five personality traits in model 2a and the indicators for resilience, overcontrolled, and undercontrolled personality types in model 2b.

Emotional Intensity

In model 1 (Table 3), it was observed that United Kingdom ($\beta_{TheNetherlands}$=–.20, P<.001), women (β_{man}=–.14, P<.001), and those imagining the unambiguous scenario ($\beta_{unambiguous}$=.12, P<.001) reported the highest emotional intensity. Model 1 accounted for 8% of the variance ($F_{4,901}$=18.28; P<.001). In model 2a (Table 3), it was observed that less emotionally stable (β=–.25, P<.001) and more agreeable (β=.12, P<.001) participants reported a higher emotional intensity. Model 2a additionally accounted for an additional 5% of the variance ($F_{9,901}$=14.64; P<.001). Model 2b (Table 3) showed that those reporting more internalizing problems (β=.33, P<.001) and more resilient participants (β=.22, P<.001) reported a higher emotional intensity. Model 2b additionally accounted for 17% of the variance ($F_{7,901}$=66.32; P<.001).

XSL•FO
RenderX

Table 3. Results of hierarchical regression analyses showing amount of variance in the emotional intensity dimension accounted for by country of residence, condition, gender, age, Big Five personality traits, DASS-21, aggression, and resilience.

Model	B[a]	SE	β[b]	*t*	*P* value	F[c]	*R*	R^2	ΔR^{2d}
1									
(Constant)	0.36	.11	—[e]	3.22	<.001	18.28	.28	.08	.08
Country[f]	–0.40	.06	–.20	–6.14	<.001	—	—	—	—
Condition[g]	0.23	.06	.12	3.60	<.001	—	—	—	—
Gender[h]	–0.28	.06	–.14	–4.34	<.001	—	—	—	—
Age	0	0	–.05	–1.39	.16	—	—	—	—
2a									
(Constant)	0.21	.11	—	1.89	.06	14.64	.36	.13	.05
Country[f]	–0.35	.06	–.18	–5.54	<.001	—	—	—	—
Condition[g]	0.21	.06	.11	3.36	<.001	—	—	—	—
Gender[h]	–0.17	.07	–.08	–2.52	.01	—	—	—	—
Age	0	0	–.02	–.540	.59	—	—	—	—
Extraversion	0.04	.05	.03	0.78	.44	—	—	—	—
Agreeableness	0.19	.06	.12	3.05	<.001	—	—	—	—
Conscientiousness	0.08	.06	.05	1.20	.23	—	—	—	—
Emotional stability	–0.37	.06	–.25	–6.59	<.001	—	—	—	—
Openness	–0.03	.07	–.01	–.390	.69	—	—	—	—
2b									
(Constant)	–2.41	.24	—	–10.27	<.001	66.32	.49	.24	.17
Country[f]	–0.29	.06	–.14	–4.85	<.001	—	—	—	—
Condition[g]	0.18	.06	.09	3.01	.003	—	—	—	—
Gender[h]	–0.32	.06	–.16	–5.42	<.001	—	—	—	—
Age	0.01	0	.14	4.21	<.001	—	—	—	—
Depression, anxiety, stress	0.44	.06	.33	7.26	<.001	—	—	—	—
Aggression	0.09	.05	.08	1.80	.07	—	—	—	—
Resilience	0.04	.01	.22	7.40	<.001	—	—	—	—

[a]B: unstandardized coefficient.

[b]β: beta standardized coefficient.

[c]F: F ratio.

[d]ΔR^2: R^2 change.

[e]Not applicable.

[f]Reference category: United Kingdom.

[g]Reference category: ambiguous situation.

[h]Reference category: women.

Proactive Versus Fight/Flight

More fight/flight reactions were reported by younger participants (β_{age}=–.26, *P*<.001), by men (β_{man}=.16, *P*<.001), and by participants responding to the unambiguous scenario ($\beta_{unambiguous}$=.09, *P*=.006). Model 1 (Table 4) accounted for 9% of the variance ($F_{4,901}$=22.59; *P*<.001). In model 2a (Table 4), it was observed that less agreeable (β=–.34, *P*<.001), less conscientious (β=–.19, *P*<.001), and less open (β=–.22, *P*<.001) but more extraverted (β=.07, *P*=.02) participants showed more fight/flight reactions. Model 2a accounted for an additional 32% of the variance ($F_{9,901}$=68.57; *P*<.001). In model 2b (Table 4), it was observed that less resilient participants (β=–.19, *P*<.001) and participants with more internalizing problems (β_{DASS}=.26, *P*<.001) and more aggression (β=.25, *P*<.001) reported more

XSL•FO
RenderX

fight/flight reactions. Model 2b accounted for an additional 24% of the variance ($F_{7,901}$=62.82; *P*<.001).

Table 4. Results of hierarchical regression analyses showing amount of variance in the proactive versus fight/flight reactions dimension accounted for by country of residence, condition, gender, age, Big Five personality traits, DASS-21, aggression, and resilience.

Model	B[a]	SE	β[b]	*t*	*P* value	F[c]	*R*	R^2	ΔR^{2d}
1									
(Constant)	0.48	.11	—[e]	4.43	<.001	22.59	.30	.09	.09
Country[f]	0	.06	0	0	>.99	—	—	—	—
Condition[g]	0.17	.06	.09	2.73	.01	—	—	—	—
Gender[h]	0.31	.06	.16	4.85	<.001	—	—	—	—
Age	–0.02	0	–.26	–8.10	<.001	—	—	—	—
2a									
(Constant)	0.08	.09	—	0.81	.42	68.57	.64	.41	.32
Country[f]	0.07	.05	.04	1.37	.17	—	—	—	—
Condition[g]	0.17	.05	.08	3.23	<.001	—	—	—	—
Gender[h]	0.07	.05	.03	1.21	.23	—	—	—	—
Age	–0.01	0	–.08	–2.92	<.001	—	—	—	—
Extraversion	0.10	.04	.07	2.44	.02	—	—	—	—
Agreeableness	–0.51	.05	–.34	–10.26	<.001	—	—	—	—
Conscientiousness	–0.31	.05	–.19	–6.08	0	—	—	—	—
Emotional stability	–0.05	.05	–.03	–0.99	.32	—	—	—	—
Openness	–0.41	.06	–.22	–7.29	<.001	—	—	—	—
2b									
(Constant)	–0.33	.22	—	–1.51	.13	62.82	.57	.33	.24
Country[f]	0.12	.06	.06	2.17	.03	—	—	—	—
Condition[g]	0.14	.06	.07	2.26	.01	—	—	—	—
Gender[h]	0.20	.06	.10	3.58	<.001	—	—	—	—
Age	–0.01	0	–.07	–2.44	.02	—	—	—	—
Depression, anxiety, stress	0.34	.06	.26	6.04	<.001	—	—	—	—
Aggression	0.27	.05	.25	5.96	<.001	—	—	—	—
Resilience	–0.04	.01	–.19	–7.02	<.001	—	—	—	—

[a]B: unstandardized coefficient.

[b]β: beta standardized coefficient.

[c]F: F ratio.

[d]ΔR^2: R^2 change.

[e]Not applicable.

[f]Reference category: United Kingdom.

[g]Reference category: ambiguous situation.

[h]Reference category: women.

Affective Versus Cognitive/Motivational

Model 1 (Table 5) showed that for older participants (β_{age}=.21, *P*<.001), men (β_{man}=.13, *P*<.001), and Dutch participants ($\beta_{TheNetherlands}$=.08, *P*=.01), the cognitive/motivational reactions were more salient. The model accounted for 7% of the variance ($F_{4,901}$=18.02; *P*<.001). Only emotional stability was a significant predictor of the salience of cognitive motivational reactions in model 2a (β=.14, *P*<.001). Model 2a (Table 5) accounted for an additional 2% of the variance ($F_{9,901}$=10.57; *P*<.001). In model 2b, it was observed that more aggression (β=.13, *P*=.01) and less internalizing problems (β=–.32, *P*<.001)

XSL•FO
RenderX

related to a comparatively higher salience of cognitive/motivational than affective reactions. Model 2b (Table 5) accounted for an additional 5% of the variance ($F_{7,901}$=18.06; P<.001).

Table 5. Results of hierarchical regression analyses showing amount of variance in the affective versus cognitive/motivational dimension accounted for by country of residence, condition, gender, age, Big Five personality traits, DASS-21, aggression, and resilience.

Model	B[a]	SE	β[b]	*t*	*P*	F[c]	*R*	R^2	ΔR^{2d}
1									
(Constant)	–0.86	.11	—[e]	–7.81	<.001	18.02	.27	.07	.07
Country[f]	0.16	.06	.08	2.52	.01	—	—	—	—
Condition[g]	0.12	.06	.06	1.79	.07	—	—	—	—
Gender[h]	0.25	.06	.13	3.95	<.001	—	—	—	—
Age	0.02	0	.21	6.59	<.001	—	—	—	—
2a									
(Constant)	–0.75	.12	—	–6.52	<.001	10.57	.31	.10	.02
Country[f]	0.11	.06	.06	1.76	.08	—	—	—	—
Condition[g]	0.13	.06	.07	2.06	.04	—	—	—	—
Gender[h]	0.23	.07	.11	3.38	<.001	—	—	—	—
Age	0.01	0	.18	5.33	<.001	—	—	—	—
Extraversion	0.06	.05	.04	1.21	.23	—	—	—	—
Agreeableness	–0.04	.06	–.03	–0.66	.51	—	—	—	—
Conscientiousness	–0.04	.06	–.02	–0.58	.56	—	—	—	—
Emotional stability	0.20	.06	.14	3.54	<.001	—	—	—	—
Openness	0.04	.07	.02	0.51	.61	—	—	—	—
2b									
(Constant)	–0.04	.25	—	–0.18	.86	18.06	.35	.12	.05
Country[f]	0.13	.06	.07	2.03	.42	—	—	—	—
Condition[g]	0.13	.06	.06	2.04	.41	—	—	—	—
Gender[h]	0.26	.06	.13	4.08	<.001	—	—	—	—
Age	0.01	0	.13	3.63	<.001	—	—	—	—
Depression, anxiety, stress	–0.42	.07	–.32	–6.50	<.001	—	—	—	—
Aggression	0.14	.05	.13	2.65	.01	—	—	—	—
Resilience	0	.01	–.01	0.32	.75	—	—	—	—

[a]B: unstandardized coefficient.

[b]β: beta standardized coefficient.

[c]F: F ratio.

[d]ΔR^2: R^2 change.

[e]Not applicable.

[f]Reference category: United Kingdom.

[g]Reference category: ambiguous situation.

[h]Reference category: women.

Discussion

Internal Structure

The first and foremost goal of this study was to investigate the structure of emotional reactions in one of the most emblematic situations of cybersecurity breaches of the upcoming IoT devices—the hacking of one's smart security camera—by looking at the full emotion process that can be elicited by this situation. Not a 1-dimensional but a 3-dimensional structure clearly emerges.

XSL•FO
RenderX

On the first dimension, all emotional reactions are loading positively. With the subjective experience items loading the highest on this dimension, this general intensity dimension can be best interpreted as a negative affectivity dimension, comparable to, for instance, the frequently used negative affectivity scale of the PANAS [13].

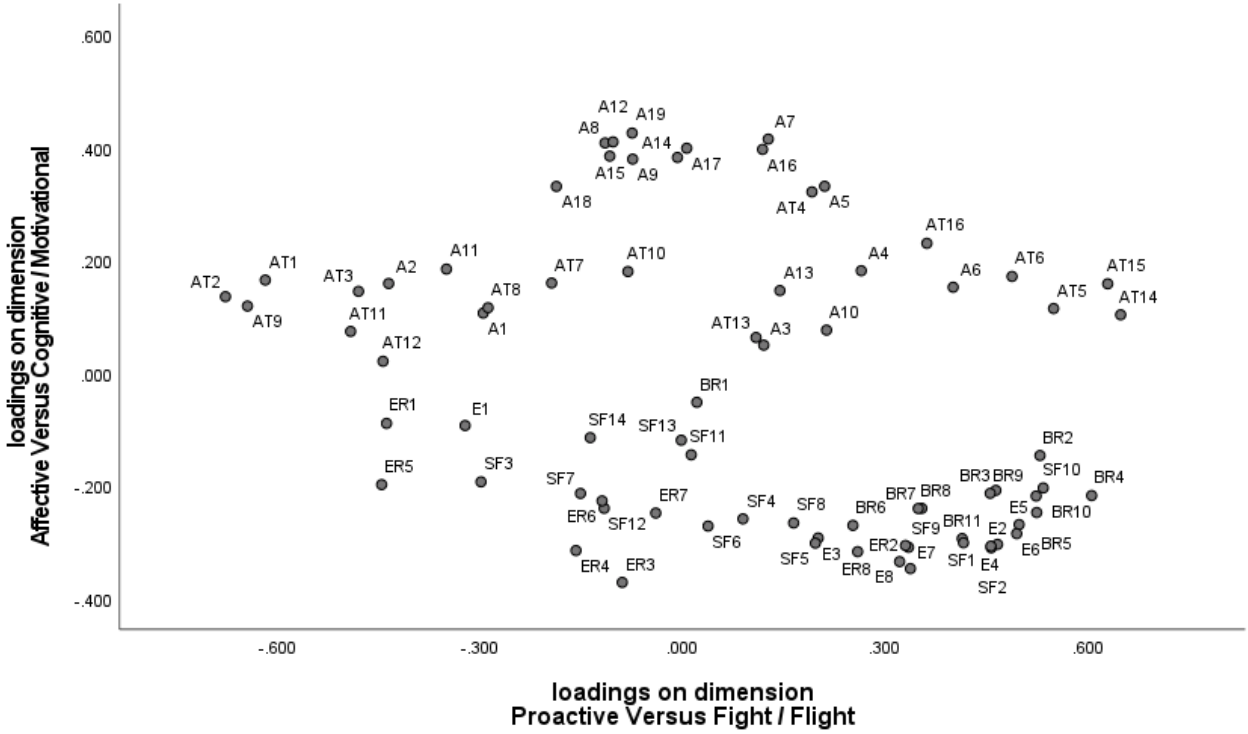
The second dimension represents the relative salience of proactive versus fight/flight action tendencies. This second dimension underlines the central status of action tendencies for the concept of emotion [44,45]. From an evolutionary perspective, emotion processes are phylogenetically shaped processes that quickly prepare the organism for action. However, depending on the concrete situation, these elicited action tendencies can be more or less constructive. In the new internet environment where we interact from a distance, acting aggressively or withdrawing are not adaptive reactions. One often does not know who is responsible and one's life depends more and more on participating in this interconnected online world. Only the proactive tendencies to stop what is happening and to better protect oneself can be considered adaptive and lead to constructive results.

The third dimension describes the relative salience of cognitive/motivational versus affective (expression, bodily reactions, regulation, and feelings) features. Possibly, this finding can be linked to the different levels of consciousness with which appraisals can occur [46]. When the appraisals are made consciously, one can focus more on what one feels inclined to do and should do. When the appraisals are made unconsciously, the way the emotion is felt and expressed becomes more salient rather than what has elicited the emotion.

When the second and third dimensions are combined, a distinction emerges that has been referred to in the stress and coping literature as problem-focused versus emotion-focused coping [47] (Figure 1). The proactive tendencies in the upper-left quadrant correspond with problem-focused coping. The bodily reactions, subjective feelings, and expressions in the lower-left quadrant indicate that one is overwhelmed and regulation is required.

This 3-dimensional structure is highly replicable: exactly the same structure was found across the two versions of the security breach scenario, across the two countries with their respective languages, and across the two genders (Multimedia Appendix 1).

Figure 1. Plot of the loadings of the emotional reactions on the second and third dimension as a function of the emotion component to which they belong. A: appraisal; AT: action tendency; BR: bodily reaction; E: expression; SF: subjective feeling; ER: emotion regulation.



Predictors of Emotion Dimensions

The second goal was to explore whether personality characteristics predict the empirically identified emotion dimensions, and, if that is the case, which ones (Tables 3-5). The general finding is that the broad personality characteristics from both personality models relate differentially to the 3 emotion dimensions, which confirms that these emotion dimensions are indeed each capturing valid aspects of the emotion processes.

Big Five Personality Model

The two most predictive personality traits are emotional stability and agreeableness. In line with the well-documented negative relationship between emotional stability and negative affectivity [22-24], we observed that emotionally stable participants scored lower on the general emotion dimension and reported a higher

XSL•FO
RenderX

salience of the affective components. Agreeable participants showed more proactive action tendencies and tended also to score a bit higher on the general emotion dimension. It is possible that agreeable people, who value warm interpersonal relationships, appraise negatively intended actions by others, like hacking, as more relevant while at the same time are less inclined to react aggressively, which frees more energy to deal constructively with the situation. Moreover, agreeable people are more likely to use cognitive restructuring and problem-solving approaches [22-24]. Conscientiousness and openness only predicted proactive tendencies. Being diligent, efficient, and orderly, which are characteristics of conscientious people, might help to focus on the action tendencies that can provide support in effectively dealing with the situation. The relationship with the personality trait openness was a bit less self-evident. As IoT is a recent and fast-developing field, people who are more curious and open are possibly more likely to understand the full implications of cybersecurity breaches and act accordingly. Extraversion, which has been found in the literature to be predictive of positive but not negative affectivity [22-24], was virtually unrelated to the emotion dimensions (with the exception of a very small although statistically significant relationship with fight/flight tendencies, probably due to the fact that extravert people tend to express their emotions more) [48].

Resilient/Overcontrolled/Undercontrolled Personality Type Model

Internalization problems, which are important characteristics of an overcontrolled personality type, predicted a higher general emotional intensity, more fight/flight tendencies, and a comparatively higher salience of the affective components. This finding indicates that people who are already vulnerable do not succeed in adequately dealing with the emotional experience. Interestingly, resilience, a characteristic of well-functioning people [43], not only predicts more proactive tendencies but also a higher general intensity of emotional reactions and a higher salience of the affective components. Possibly because resilient people can cope better with stressors, they are less defensive and more willing to appraise the seriousness of the situation and accept their own emotional reactions. Finally, aggression, as an indicator of an undercontrolled personality type, is especially predictive of fight/flight action tendencies and relates to a slightly higher salience of the cognitive/motivational components. People who are high on aggression are more willing to blame others and are primed on aggressive reactions [49,50].

Gender and Age

In addition to the personality predictors, we also found that gender and age played a role. Women had a tendency to have more emotionally intense and affective reactions, while men were more likely to show fight/flight reactions. This is in line with earlier findings that females generally have more intense emotional reactions [25-28], experience more emotions in situations of cyberbullying [51], and have more anxiety in situations of hacking [52] and that males tend to react more aggressively [53].

We also found that older individuals were more prone to have proactive and cognitive/motivational reactions, which fits the observation that older individuals have less negative affective experiences and healthier emotion regulation strategies [29-32]. Additionally, intra-individual differences in emotional reactions to cybersecurity breaches are organized and structured in exactly the same way for males and females (Multimedia Appendix 1).

Ambiguous/Nonambiguous Conditions

While the two cybersecurity scenarios showed exactly the same 3-dimensional structure of emotional reactions, quantitative differences were observed, with the nonambiguous situation eliciting more intense and more fight/flight emotional reactions. The nonambiguous situation is possibly experienced more as though one is confronted with a natural person in real life. The situation becomes more relevant for one's goals and elicits more fight/flight action tendencies rather than the more adaptive proactive reactions.

Country

Finally, while the emotion structure is the same in the two countries, we observed less intense reactions and a higher salience of cognitive/motivational reactions in the Dutch as compared with the UK sample. A speculative explanation could be that the emblematic example of the hacking of a smart security camera has received more media coverage in the United Kingdom [54-57] than in the Netherlands, which has made these scenarios more emotionally salient in the United Kingdom.

Principal Findings

In this study, only the direct emotional reactions to a cybersecurity breach scenario have been studied. A question for future research is whether and to what extent these immediate emotional reactions set the stage for further mental health problems. Being exposed to hacking has been linked to psychopathology [58-61] in the literature and even to suicide in the media [9]. Based on the discovered emotion structure, very different dynamics can be predicted with respect to the role that a hacking situation will play in a person's life in the longer run.

Those with intense emotional reactions, fight/flight action tendencies, and salient affective components are probably more likely to stay confronted with the situation and its negative ramifications. They experience the situation as emotionally highly relevant, but they tend to react in a way that does not resolve the challenges created by the problem. Moreover, they are additionally confronted with affective reactions that need to be regulated and thus require extra energy. This combination can be considered the psychologically least adaptive reaction, which sets the stage for further mental health complaints.

Those who have no or little negative emotional reactions can only be partially considered better off. They do not have to deal with the negative emotional reactions themselves but also lack the inherent pressure created in the emotion process to take action. Emotions are relevance detectors [5]. Appraising the situation as threatening with its ensuing negative emotional reactions can motivate appropriate action and can therefore be considered adaptive. This interpretation is also supported by

XSL•FO
RenderX

the finding that resilient people score higher on the general intensity dimension.

The most adaptive emotional reaction can be considered to be a negative emotional reaction in which the proactive and constructive action tendencies and cognitive-motivational components are the most salient. Such a reaction pattern implies that the seriousness of the situation is adequately appraised and thus that the emotions play their role as relevance detectors. At the same time, actions are prepared that maximize an effective resolution of the situation without the person being overwhelmed by the affective reactions.

Limitations

One of the limitations of this study is that the causal conclusions about the long-term mental health consequences of a cybersecurity breach cannot be investigated with a scenario methodology based on anticipated emotional experiences. However, as experimental research of real emotional experiences is impossible or at least highly limited in this area due to ethical considerations (it is unethical to actually invade the privacy of people by hacking their security camera), scenarios offer an ethically viable and direct way to study the structure of emotional reactions in this uncharted domain. As this study was conducted in Western Europe, further cultural generalizability is yet to be demonstrated. Future research can also study the ecological validity, generalizability, and long-term mental health implications of these findings. Another limitation is the use of self-assessment instruments. While some emotion components can only be studied through self-assessment (like subjective feelings and cognitive appraisals), other components can be studied by objective data (like psychophysiological and expressive changes). In future research, it would be interesting to complement self-reported data with such objective data.

Conclusion

With the increasing interconnections through the internet and especially the recent development of IoT, people are much more at risk of experiencing cybersecurity breaches. Becoming a victim of cybersecurity breaches, with possibly far-reaching consequences for one's personal and professional life, is becoming more and more likely. When all components of the emotion processes elicited by such cybersecurity breaches are investigated, a replicable 3-dimensional structure emerges that goes beyond the well-known negative affectivity dimension. These dimensions relate differentially to broad personality characteristics, which further validates the need for a multidimensional representation. Depending on the position of the emotional reaction on these three dimensions, very different predictions can be made about the long-term mental health implications of hacking experiences. With this study, a key process that links the occurrence of a cybersecurity breach situation with possible long-term mental health effects has been mapped out.

Acknowledgments

This study is part of a larger research project, Emotion Psychology Meets Cyber Security in IoT Smart Homes (Cocoon), funded by EU FP7 CHIST-ERA funding scheme (European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-NET) corresponding to grants FWO project G0H6416N-FWOOPR2016009701 and EPSRC EP/P016448/1 and NWO project no.651.002.002. Many thanks to native speaker Arpine Hovasapian, PhD, for proofreading the paper. We thank her for her valuable contribution to the manuscript.

Authors' Contributions

SB and JRJF substantially contributed to analyses and interpretation of data, drafted the article, and gave final approval of the version to be published. SB, JRJF, NMAH, AH, GL, and EBR substantially contributed to conception of the study and design and acquisition of data, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplement.
[PDF File (Adobe PDF File), 205 KB - jmir_v23i5e24879_app1.pdf]

Multimedia Appendix 2
Cybersecurity GRID.
[PDF File (Adobe PDF File), 83 KB - jmir_v23i5e24879_app2.pdf]

References

1. Security Breach. 2019. URL: https://www.techopedia.com/definition/29060/security-breach [accessed 2021-03-30]
2. Threat Report: Protecting people—a quarterly analysis of highly targeted attacks. Proofpoint. 2018. URL: https://www.proofpoint.com/au/resources/threat-reports/cybersecurity-analysis-report [accessed 2021-03-30]

XSL•FO
RenderX

3. The most telling cyber security statistics in 2019. Techjury. URL: https://techjury.net/stats-about/cyber-security/ [accessed 2021-04-26]
4. Frijda NH. The Emotions. Cambridge: Cambridge University Press; 1986.
5. Scherer KR. Appraisal considered as a process of multilevel sequential checking. In: Appraisal Processes in Emotion: Theory, Methods, Research. New York: Oxford University Press; 2001.
6. The Norton Online Family Report. Symantec. 2010. URL: https://www.broadcom.com/404-symantec [accessed 2018-03-30]
7. Beaton A, Cook M, Kavanagh M, Herrington C. The psychological impact of burglary. Psychol Crime Law 2008 Jan 04;6(1):33-43. [doi: 10.1080/10683160008410830]
8. Chung MC, Stedmon J, Hall R, Marks Z, Thornhill K, Mehrshahi R. Posttraumatic stress reactions following burglary: the role of coping and personality. Traumatol Int J 2014 Jun;20(2):65-74. [doi: 10.1037/h0099374]
9. Baraniuk C. Ashley Madison ″suicides″ over website hack. BBC News. 2015 Aug 24. URL: https://www.bbc.com/news/technology-34044506 [accessed 2019-12-10]
10. Gaps in European cyber education and professional training: education, training, awareness, cyber ranges. European Cyber Security Organisation. 2017. URL: https://ecs-org.eu/documents/publications/5fdb282a4dcbd.pdf [accessed 2021-03-30]
11. Vigdor N. Somebody's watching: hackers breach ring home security cameras. The New York Times. 2019 Dec 15. URL: https://www.nytimes.com/2019/12/15/us/Hacked-ring-home-security-cameras.html [accessed 2021-03-30]
12. Buntz B. 5 Cybersecurity lessons related to IP security cameras. IoT World Today. 2019 Aug 31. URL: https://www.iotworldtoday.com/2019/08/31/5-cybersecurity-lessons-related-to-ip-security-cameras/ [accessed 2021-03-30]
13. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Personality Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
14. Fontaine JRJ, Scherer KR, Soriano C. Components of Emotional Meaning: A Sourcebook. Oxford: Oxford University Press; 2013.
15. Frijda NH. Impulsive action and motivation. Biol Psychol 2010 Jul;84(3):570-579. [doi: 10.1016/j.biopsycho.2010.01.005] [Medline: 20064583]
16. Fontaine JRJ, Scherer KR, Roesch EB, Ellsworth PC. The world of emotions is not two-dimensional. Psychol Sci 2007 Dec;18(12):1050-1057. [doi: 10.1111/j.1467-9280.2007.02024.x] [Medline: 18031411]
17. Goldberg LR. The development of markers for the Big Five factor structure. Psychol Assess 1992;4(1):26-42. [doi: 10.1037/1040-3590.4.1.26]
18. Robins RW, John OP, Caspi A, Moffitt TE, Stouthamer-Loeber M. Resilient, overcontrolled, and undercontrolled boys: three replicable personality types. J Pers Soc Psychol 1996 Jan;70(1):157-171. [doi: 10.1037//0022-3514.70.1.157] [Medline: 8558407]
19. Donnellan MB, Robins RW. Resilient, overcontrolled, and undercontrolled personality types: issues and controversies. Soc Personality Psychol Compass 2010;4(11):1070-1083. [doi: 10.1111/j.1751-9004.2010.00313.x]
20. Schindler S, Querengässer J. Coping with sadness—how personality and emotion regulation strategies differentially predict the experience of induced emotions. Personality Indiv Diff 2019 Jan;136:90-95. [doi: 10.1016/j.paid.2018.01.050]
21. Sander D, Scherer KR. The Oxford Companion to Emotion and the Affective Sciences. Oxford: Oxford University Press; 2009.
22. Carver CS, Connor-Smith J. Personality and coping. Annu Rev Psychol 2010;61:679-704. [doi: 10.1146/annurev.psych.093008.100352] [Medline: 19572784]
23. 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]
24. Hughes DJ, Kratsiotis IK, Niven K, Holman D. Personality traits and emotion regulation: a targeted review and recommendations. Emotion 2020 Feb;20(1):63-67. [doi: 10.1037/emo0000644] [Medline: 31961180]
25. Bradley MM, Codispoti M, Sabatinelli D, Lang PJ. Emotion and motivation II: sex differences in picture processing. Emotion 2001;1(3):300-319. [doi: 10.1037/1528-3542.1.3.300]
26. Brebner J. Gender and emotions. Personality Indiv Diff 2003 Feb;34(3):387-394. [doi: 10.1016/s0191-8869(02)00059-4]
27. Deng Y, Chang L, Yang M, Huo M, Zhou R. Gender differences in emotional response: inconsistency between experience and expressivity. PLoS One 2016 Jun 30;11(6):e0158666 [FREE Full text] [doi: 10.1371/journal.pone.0158666] [Medline: 27362361]
28. Fischer AH, Rodriguez Mosquera PM, van Vianen AEM, Manstead ASR. Gender and culture differences in emotion. Emotion 2004 Mar;4(1):87-94. [doi: 10.1037/1528-3542.4.1.87] [Medline: 15053728]
29. Mroczek DK. Age and emotion in adulthood. Curr Dir Psychol Sci 2016 Jun 24;10(3):87-90. [doi: 10.1111/1467-8721.00122]
30. John OP, Gross JJ. Healthy and unhealthy emotion regulation: personality processes, individual differences, and life span development. J Pers 2004 Dec;72(6):1301-1333. [doi: 10.1111/j.1467-6494.2004.00298.x] [Medline: 15509284]
31. Johnson SJ, Machowski S, Holdsworth L, Kern M, Zapf D. Age, emotion regulation strategies, burnout, and engagement in the service sector: advantages of older workers. Rev Trabajo y las Organizaciones 2017;33(3):205-216. [doi: 10.1016/j.rpto.2017.09.001]
32. Hay EL, Diehl M. Emotion complexity and emotion regulation across adulthood. Eur J Ageing 2011 Sep 17;8(3):157-168 [FREE Full text] [doi: 10.1007/s10433-011-0191-7] [Medline: 21941465]

XSL•FO
RenderX

33. Świątkowski W, Dompnier B. Replicability crisis in social psychology: looking at the past to find new pathways for the future. Int Rev Soc Psychol 2017 May 02;30(1):111. [doi: 10.5334/irsp.66]
34. Roberts C, Gilbert E, Allum N, Eisner L. Research synthesis. Public Opin Q 2019 Nov 12;83(3):598-626. [doi: 10.1093/poq/nfz035]
35. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;140:1-55.
36. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]
37. Bryant FB, Smith BD. Refining the architecture of aggression: a measurement model for the Buss-Perry Aggression Questionnaire. J Res Personality 2001 Jun;35(2):138-167. [doi: 10.1006/jrpe.2000.2302]
38. Diamond PM, Wang EW, Buffington-Vollum J. Factor structure of the Buss-Perry Aggression Questionnaire (BPAQ) with mentally ill male prisoners. Crim Justice Behav 2016 Jun 30;32(5):546-564. [doi: 10.1177/0093854805278416]
39. Diamond PM, Magaletta PR. The short-form Buss-Perry Aggression Questionnaire (BPAQ-SF): a validation study with federal offenders. Assessment 2006 Sep 26;13(3):227-240. [doi: 10.1177/1073191106287666] [Medline: 16880276]
40. Buss AH, Perry M. The Aggression Questionnaire. J Personality Soc Psychol 1992 Sep;63(3):452-459. [doi: 10.1037/0022-3514.63.3.452]
41. Alessandri G, Vecchione M, Caprara G, Letzring TD. The Ego Resiliency Scale Revised. Eur J Psychol Assess 2012 Nov;28(2):139-146. [doi: 10.1027/1015-5759/a000102]
42. Vecchione M, Alessandri G, Barbaranelli C, Gerbino M. Stability and change of ego resiliency from late adolescence to young adulthood: a multiperspective study using the ER89-R Scale. J Pers Assess 2010 May 19;92(3):212-221. [doi: 10.1080/00223891003670166] [Medline: 20408021]
43. Block J, Kremen AM. IQ and ego-resiliency: conceptual and empirical connections and separateness. J Personality Soc Psychol 1996;70(2):349-361. [doi: 10.1037/0022-3514.70.2.349]
44. Frijda NH, Kuipers P, ter Schure E. Relations among emotion, appraisal, and emotional action readiness. J Personality Soc Psychol 1989;57(2):212-228. [doi: 10.1037/0022-3514.57.2.212]
45. Frijda NH. Emotion experience. Cogn Emotion 2005 Jun;19(4):473-497. [doi: 10.1080/02699930441000346]
46. Leventhal H, Scherer KR. The relationship of emotion to cognition: a functional approach to a semantic controversy. Cogn Emotion 1987 Mar;1(1):3-28. [doi: 10.1080/02699938708408361]
47. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer Publishing Company; 1984.
48. Riggio HR, Riggio RE. Emotional expressiveness, extraversion, and neuroticism: a meta-analysis. J Nonverbal Behav 2002;26(4):195-218. [doi: 10.1023/A:1022117500440]
49. Elison J, Garofalo C, Velotti P. Shame and aggression: theoretical considerations. Aggress Violent Behav 2014 Jul;19(4):447-453. [doi: 10.1016/j.avb.2014.05.002]
50. Zimmer-Gembeck MJ, Nesdale D, Webb HJ, Khatibi M, Downey G. A longitudinal rejection sensitivity model of depression and aggression: unique roles of anxiety, anger, blame, withdrawal and retribution. J Abnorm Child Psychol 2016 Oct 27;44(7):1291-1307. [doi: 10.1007/s10802-016-0127-y] [Medline: 26816212]
51. Balakrishnan V. Actions, emotional reactions and cyberbullying—from the lens of bullies, victims, bully-victims and bystanders among Malaysian young adults. Telematics Informat 2018 Aug;35(5):1190-1200. [doi: 10.1016/j.tele.2018.02.002]
52. Elhai JD, Chai S, Amialchuk A, Hall BJ. Cross-cultural and gender associations with anxiety about electronic data hacking. Comput Human Behav 2017 May;70:161-167. [doi: 10.1016/j.chb.2017.01.002]
53. Bettencourt BA, Miller N. Gender differences in aggression as a function of provocation: a meta-analysis. Psychol Bull 1996 May;119(3):422-447. [doi: 10.1037/0033-2909.119.3.422] [Medline: 8668747]
54. Gibbs S. The Guardian. 2014 Nov 20. URL: http://www.theguardian.com/technology/2014/nov/20/webcam-russians [accessed 2020-10-09]
55. Noor P. Ring hackers are reportedly watching and talking to strangers via in-home cameras. The Guardian. 2019 Dec 13. URL: http://www.theguardian.com/technology/2019/dec/13/ring-hackers-reportedly-watching-talking-strangers-in-home-cameras [accessed 2020-10-09]
56. Paul K. Ring sued by man who claims camera was hacked and used to harass his kids. The Guardian. 2019 Dec 27. URL: http://www.theguardian.com/technology/2019/dec/27/ring-camera-lawsuit-hackers-alabama [accessed 2020-10-09]
57. Weaver M. UK moves to shut down Russian hackers streaming live British webcam footage. The Guardian. 2014 Nov 20. URL: http://www.theguardian.com/technology/2014/nov/20/webcam-hackers-watching-you-watchdog-warns [accessed 2020-10-09]
58. Kopecký K, Szotkowski R. Cyberbullying, cyber aggression and their impact on the victim—the teacher. Telematics Informat 2017 May;34(2):506-517. [doi: 10.1016/j.tele.2016.08.014]
59. Gross ML, Canetti D, Vashdi DR. The psychological effects of cyber terrorism. Bull At Sci 2016 Aug 04;72(5):284-291 [FREE Full text] [doi: 10.1080/00963402.2016.1216502] [Medline: 28366962]
60. Gross ML, Canetti D, Vashdi DR. Cyberterrorism: its effects on psychological well-being, public confidence and political attitudes. J Cybersecurity 2017;3(1):49-58. [doi: 10.1093/cybsec/tyw018]

XSL•FO
RenderX

61. Canetti D, Gross M, Waismel-Manor I, Levanon A, Cohen H. How cyberattacks terrorize: cortisol and personal insecurity jump in the wake of cyberattacks. Cyberpsychol Behav Soc Netw 2017 Feb;20(2):72-77. [doi: 10.1089/cyber.2016.0338] [Medline: 28121462]

Abbreviations

DASS-21: Depression, Anxiety and Stress Scale–21 Item
IoT: Internet of Things
IPIP-50: International Personality Item Pool 50
PANAS: positive and negative affect schedule

Edited by R Kukafka; submitted 09.10.20; peer-reviewed by A Teles, V Franzoni; comments to author 14.12.20; revised version received 05.02.21; accepted 16.03.21; published 12.05.21.

Please cite as:
Budimir S, Fontaine JRJ, Huijts NMA, Haans A, Loukas G, Roesch EB
Emotional Reactions to Cybersecurity Breach Situations: Scenario-Based Survey Study
J Med Internet Res 2021;23(5):e24879
URL: https://www.jmir.org/2021/5/e24879
doi:10.2196/24879
PMID:33978591

©Sanja Budimir, Johnny R J Fontaine, Nicole M A Huijts, Antal Haans, George Loukas, Etienne B Roesch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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 Barriers to Novel Data Linkages: Topic Modeling of the Results of the LifeInfo Survey

Holly Clarke[1], MA; Stephen Clark[1,2], PhD; Mark Birkin[1,2], PhD; Heather Iles-Smith[3,4], PhD; Adam Glaser[1,3,5], DM; Michelle A Morris[1,5], PhD

[1]Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom

[2]School of Geography, University of Leeds, Leeds, United Kingdom

[3]Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

[4]School of Health and Society, University of Salford, Salford, United Kingdom

[5]Leeds Institute of Medical Research, School of Medicine, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Michelle A Morris, PhD
Leeds Institute for Data Analytics
University of Leeds
Level 11, Worsley Building
Claredon Way
Leeds, LS2 9NL
United Kingdom
Phone: 44 0113 343 9680
Email: M.Morris@leeds.ac.uk

Abstract

Background: Novel consumer and lifestyle data, such as those collected by supermarket loyalty cards or mobile phone exercise tracking apps, offer numerous benefits for researchers seeking to understand diet- and exercise-related risk factors for diseases. However, limited research has addressed public attitudes toward linking these data with individual health records for research purposes. Data linkage, combining data from multiple sources, provides the opportunity to enhance preexisting data sets to gain new insights.

Objective: The aim of this study is to identify key barriers to data linkage and recommend safeguards and procedures that would encourage individuals to share such data for potential future research.

Methods: The *LifeInfo Survey* consulted the public on their attitudes toward sharing consumer and lifestyle data for research purposes. Where barriers to data sharing existed, participants provided unstructured survey responses detailing what would make them more likely to share data for linkage with their health records in the future. The topic modeling technique latent Dirichlet allocation was used to analyze these textual responses to uncover common thematic topics within the texts.

Results: Participants provided responses related to sharing their store loyalty card data (n=2338) and health and fitness app data (n=1531). Key barriers to data sharing identified through topic modeling included data safety and security, personal privacy, requirements of further information, fear of data being accessed by others, problems with data accuracy, not understanding the reason for data linkage, and not using services that produce these data. We provide recommendations for addressing these issues to establish the best practice for future researchers interested in using these data.

Conclusions: This study formulates a large-scale consultation of public attitudes toward this kind of data linkage, which is an important first step in understanding and addressing barriers to participation in research using novel consumer and lifestyle data.

(J Med Internet Res 2021;23(5):e24236) doi:10.2196/24236

KEYWORDS

topic modeling; text analysis; lifestyle data; consumer data; mHealth; loyalty card; fitness tracker; data linkage; data sharing; public attitudes; public opinion

XSL•FO
RenderX

Introduction

Background

Poor diet and physical inactivity are known to contribute to millions of early deaths worldwide [1,2]. In the United Kingdom, 1 in 7 deaths are attributed to poor diet, whereas 1 in 6 deaths are attributed to physical inactivity [3,4]. A greater understanding of these risk factors for lifestyle-influenced diseases such as type 2 diabetes, certain cancers, and cardiovascular diseases is needed to improve global health. At the same time, technological advancements have led to increasingly large volumes of *big data* being produced about individual food consumption and exercise habits [5,6].

Historically, a major barrier to research on lifestyle risk factors for noncommunicable diseases has been the availability of accurate, robust, and reproducible data on diet and exercise [7,8]. Big and novel lifestyle data, produced when using services such as supermarket loyalty cards or health and fitness monitoring apps, have many benefits compared with more traditional forms of data collected through surveys, interviews, and food or exercise logs; as these data are collected during everyday activities, they are naturalistic and nonintrusive [9], meaning they do not encounter the selective reporting bias entailed with traditional methods [7]. Furthermore, large volumes of data can potentially be shared with researchers almost in real time, surpassing the scale of traditional methods at a very low cost and requiring little or no effort on the part of the participant [10]. Consequently, these data are uniquely set up for at-scale longitudinal studies with the additional benefit of extending research into traditionally hard-to-reach populations [10].

These data include mobile phone step counts, GPS-tracked exercise, wearable device heart rate monitoring, and store loyalty card records. In health research, few studies have demonstrated the full utility of consumer and personal data of this sort, as they have not typically been available to researchers [7,10,11]. Nonetheless, initiatives such as the Consumer Data Research Center [12] have begun to facilitate access to novel data sources.

In the context of diet and health research, attention has particularly been drawn to the potential of using supermarket loyalty card data (eg, Tesco Club Card) to understand food and drink purchase behavior [7,10,11,13-16] and data from wearable devices (eg, Fitbit and Garmin) or mobile phone fitness apps (eg, MyFitnessPal and Strava) to understand exercise behavior [17-19]. However, studies that have used commercial lifestyle data for research purposes have reported low uptake [10]. Although this may have been influenced by factors such as the methods used to contact participants, there is a clear research need to understand participants' reluctance to share their data.

The combination of data from multiple sources to create enhanced data sets, known as *data linkage*, provides new insights for health research that surpass those provided by the data sets individually. The value of consumer and lifestyle data is further amplified when combined with health outcomes data [10]; for example, Aiello et al [11] used supermarket loyalty card data for small geographic areas to study the association between food purchasing and health outcomes. We believe that similar work linking individual health outcomes and lifestyle data, rather than at the ecological study level, would provide added benefits through greater specificity and personalization. However, an individualized approach may highlight data privacy and ethics barriers, especially in light of understandable historical concerns regarding data linkages, such as those proposed by care.data in the United Kingdom in 2013 and 2016 [20]. In addition, data linkage can create disclosure concerns or can increase sensitivity, which must be addressed by researchers seeking to use such methods [21,22].

Public attitudes toward health data sharing for research appear to be dependent on many factors, with the most prominent being the actual data sharing process, including data security considerations [23], the purpose and social license for the research [23-25], and the level of sensitivity attributed to data [26]. Others have reported high levels of trust in health institutions, lower levels of trust in academics, and the lowest levels of trust in private companies for data sharing initiatives [24,27]. As research using big lifestyle data often involves all 3 of these factors, it is important to address how trust might influence people's willingness to participate in research under these circumstances.

Objectives

Despite the obvious opportunities provided by the proliferation of big data for health research, little is known about public attitudes toward the linkage of lifestyle data with individual health records for research. The *LifeInfo Survey* is the first of its kind at this scale (n=7101 participants) to consult the public on their attitudes toward sharing novel forms of consumer and lifestyle data for linkage with their health records for health research [28]. Moreover, this survey included free-text response questions that allowed individuals, in their own words, to state what actions would alleviate their concerns about data sharing in this setting. Surveys do not frequently allow for unstructured answers of this kind because of the subjectivity and time commitment imposed with the qualitative coding of texts [29]. The use of novel data science methods, including topic modeling, can facilitate the semiautomated analysis of large amounts of textual data to identify latent themes [30].

This study aims to advance the understanding of attitudes toward data sharing by identifying specific barriers that present themselves when linking store loyalty cards or health and fitness app data with individual health records for research purposes. We recommend procedures and safeguards that can be applied to future research linking lifestyle and health data to increase participant support.

We hypothesize that common issues identified in the literature on public attitudes toward data sharing for health research, such as data security and trust, will be evident within survey responses in addition to concerns specific to the type of data to be linked, in this case, store loyalty card and health and fitness app data.

XSL•FO
RenderX

Methods

Data Collection

To gather public opinion on data sharing and linkage, the *LifeInfo Survey* recruited participants between September 2017 and October 2019 across 2 health settings, the Leeds Teaching Hospitals National Health Service Trust and Low Moor Medical General Practice Surgery, and 2 nonhealth settings, the Leeds Institute for Data Analytics and the Leeds City Council Health Communities survey [31]. The *LifeInfo Survey* (Multimedia Appendix 1) addressed a hypothetical scenario—whether, if asked for a future study, respondents would give permission for their consumer and lifestyle data to be linked with their health records for health research. This was conditional on their data being stored safely and not shared with anyone outside the research team. The survey consulted participants specifically about 2 types of data: (1) consumer data from store loyalty cards detailing food and drink purchases and (2) lifestyle data from health and fitness apps, websites, and wearable devices. Basic demographic data and which specific loyalty cards and health and fitness apps respondents used were additionally captured. Additional information about the *LifeInfo* project can be found in the study protocol [32]. Detailed information on the participants and the main survey results are reported elsewhere [28].

Those who responded *no* or *not sure* to whether they would share their data were asked, "what (if anything) might make you change your mind in the future?" concerning their (1) store loyalty card data and (2) health and fitness app data. This study primarily analyzes the qualitative responses to these 2 questions (questions 4 and 9 in the original questionnaire included in Multimedia Appendix 1), henceforth referred to as (1) *the store loyalty card question* and (2) *the health and fitness app question*, to answer the following research question: "What are the reported barriers to linkage of lifestyle data with health records for research?" Responses were given in free-text format, allowing individuals to state, in their own words, potential desired changes that would make them more willing to share such data, although many also used this space to explain their reasons behind more negative responses. Primary analysis regarding overall willingness to share lifestyle and data demographic trends are reported elsewhere [28] and summarized below. The survey questionnaire and all responses were provided in English.

Ethics

This study was granted ethical approval by the London-Brent Research Ethics Committee (reference 17/LO/0622).

Modeling

Latent Dirichlet allocation (LDA) was applied as a method of automated content analysis on unstructured survey responses. This technique was used to identify the underlying factors that contribute to respondents' unwillingness or unsureness to having their consumer or lifestyle data and health records linked for research purposes and potential changes that could influence them to do so. LDA is a generative probabilistic model that is frequently applied to textual data. The model has a 3-level hierarchical Bayesian structure under which each *document* is modeled as several topics, and each topic is modeled as a set of terms [30]. The model uses the Gibbs sampling technique to estimate model parameters. The LDA modeling procedure was applied to free-text responses separately for the store loyalty card question and the health and fitness app question to create a model for each.

Processing

Data cleaning and processing were performed using the R software [33]. Noninformative responses (eg, *N/A* and *no comment*) were removed from the data set. Survey responses that consisted of only a single word were removed from the analysis data set, as the underlying mechanisms of LDA are based on the co-occurrence of terms. These are categorized separately, as attitudes are easily ascertained from single-word responses (shown in Tables 1 and 2).

Table 1. Counts of single-word responses to the question, "What (if anything) might make you change your mind in the future?" about sharing store loyalty card data for linkage with health records (n=396).

Single word	Counts per word, n (%)
Nothing	256 (64.6)
No	98 (24.8)
Privacy	10 (2.5)
None	8 (2.0)
Maybe	5 (1.2)
Private	3 (0.8)
Confidentiality; personal	2 (0.5)
Confidential; discounts; dk[a]; hackers; illegible; incentives; money; unneeded; unlikely; unsure; why; yes	1 (0.3)

[a]dk: don't know.

XSL•FO
RenderX

Table 2. Counts of single-word responses to the question, "What (if anything) might make you change your mind in the future?" about sharing fitness app data for linkage with health records (n=309).

Single word	Count per word, n (%)
Nothing	189 (61.2)
No	71 (23.0)
Privacy; security	7 (2.3)
Maybe; none	5 (1.6)
q4[a]	4 (1.3)
Anonymity; confidential; private; same; yes	2 (0.6)
Benefit; confidentiality; intrusive; might; nil; personal; possibly; relevance; uncertain; unlikely; why	1 (0.3)

[a]q4: question 4.

Preprocessing procedures, standard in natural language processing, were undertaken to create a *document term matrix* (DTM) on which to perform LDA. This included converting all words to lower case and removing white spaces, punctuations, and common *stop-words* from texts to leave only meaningful words. Frequent misspellings were replaced with their correct form, and common equivalent words were standardized (Multimedia Appendix 2).

The *SMART* (System for the Mechanical Analysis and Retrieval of Text) stop-word data set was used to identify uninformative words. However, bigrams (two-word terms) that contain many common stop-words were not removed to uncover attitude positions within responses (Multimedia Appendix 2). Numbers were not removed as numeric bigrams are potentially meaningful (eg, *100 percent* and *3rd party*). Finally, lemmatization of words was undertaken to convert words into their root form (*lemma*) and reduce sparsity within the DTM (eg, *cards* was replaced with *card*; Multimedia Appendix 2). Lemmatization was preferred over stemming, as lemmas are more human-readable than stems, which are not always complete words (eg, *storing* would be replaced with the stem *stor*).

LDA was performed on a DTM of unigrams (one-word terms) and bigrams (two-word terms) within the texts. Bigrams are included in the DTM as this creates more human-interpretable topics, and many words are context-specific, for example, *big brother*/*change mind*, or formulate attitude positions in combination, for example, *would change*/*wouldn't change*. In the example given, the term *brother* alone would provide little insight into data sharing attitudes, yet *big brother* signifies a potential invasion of privacy and distrust. Several papers have reported improved results for including bigrams and higher n-grams in different topic models [34-36]. Only terms that occurred in more than 2 documents were retained, leaving 993 unique terms for the store loyalty card question and 554 unique terms for the health and fitness app question.

Topic Number Selection

Selecting an appropriate number of topics is a key challenge for LDA. According to Green et al [37], "too few topics will produce results that are overly broad, whereas choosing too many will result in the 'over-clustering' of a corpus into many small, highly-similar topics" that are difficult to interpret in a meaningful way. The number of topics (k) is conventionally chosen as the model with the lowest value of *perplexity* when applying different models of candidate k to held-out data [30,38]. This perplexity measure captures how well a probability distribution or probability model predicts a sample, indicating how *surprised* the model is by new data. For LDA, it is equivalent to the inverse of the geometric mean per-word likelihood calculated on the held-out data [30]. For this analysis, ten-fold crossvalidation was used to select an appropriate number of topics (k) from a candidate list of 15 k ranging from 2 to 100, optimizing for perplexity. Candidate k increases in smaller intervals between lower values, as greater change is expected between these values. The crossvalidation process randomly divides the data set into 10 approximately equally sized folds and uses 9 of these to train the model, using the held-out fold to test the model. This process was repeated 10 times such that each fold was used as the testing set once.

Some research has found that the models that produce the most semantically meaningful topics—in that topics are easily interpreted by humans and terms representing concepts are given high probabilities within the model—are not necessarily the models with the best perplexity scores [39]. Hence, a measurement of *coherence*, which research finds corresponds well with human-interpretable topics [40], was also considered. Average topic *probabilistic coherence* measures topic quality based on how commonly topic terms co-occur, controlling for statistical independence [41-43]. This was compared for models of candidate k topics and balanced with perplexity scores.

Once an appropriate number of topics were selected for each LDA model, the final models were created using responses from all individuals for the given question. This method was chosen, rather than using commonly used training and testing approaches, to support the study aim of summarizing survey responses rather than creating a predictive model to categorize new data, as no new *LifeInfo Survey* data will be collected in the future. Moreover, the size of the data set is small compared with many others that use LDA, and splitting the data set into fewer responses could reduce the model quality.

For both topic number selection and the final models, the LDA hyperparameters were set at α=.1, influencing document-topic density, and β=.05, influencing word-topic density. The α prior was set at this relatively low value because *LifeInfo Survey* responses were short (refer to Multimedia Appendix 3 for plots

XSL•FO
RenderX

showing the distribution of response lengths in words, averaging 12 words per document for the store loyalty card question and 9 words for the health and fitness app question), so we would expect there to be only a few topics formulating each document. β was also set to a relatively low value, as we expected a small number of words to be highly influential per topic given the short responses. α is modeled as asymmetrical, as we expected some topics to be more common than others within the survey responses. Previous work has found that asymmetrical α values provide substantial advantages to LDA results, whereas asymmetrical β priors provide no benefit [44].

Hierarchical Clustering

To make the topics more easily interpretable, those created through LDA modeling were further categorized into thematic groups with the aid of hierarchical clustering of topics (Multimedia Appendix 4). These topics were given summarizing names, shown in Tables 3 and 4 by subheadings in bold, based on their content, considering both topic *top terms* and the contextual use of these terms in texts (Multimedia Appendix 5). The Hellinger distance [45] between topics was calculated based on their term φ values, and the 2 closest topics were clustered together. φ values are the probability of a term being used within a text given; therefore, topics that more frequently use the same terms are considered closer. This was done iteratively until hierarchical clusters split topics into meaningful thematic categories. In some cases, topics thematically aligned with a category that was found to be semantically different to the topic according to hierarchical clustering; these are marked with a superscript in the tables and dendrogram and were reassigned to their appropriate thematic category.

XSL•FO
RenderX

Table 3. Topics created by latent Dirichlet allocation modeling of LifeInfo store loyalty card question, showing topic names, top 15 terms per topic according to term φ values, topic prevalence, and topic probabilistic coherence. In total, 9 thematic categories are shown.

Themes and SC[a] topic number	Topic name: top 15 terms selected by highest probability of the term given the topic	Prevalence (estimated survey responses; n=1930)[b], n (%)	Coherence
Nothing would change mind			
SC14	Wouldn't change mind: change, mind, change mind, nothing, would change, nothing would, don't think, make, wouldn't change, nothing change, think would, anything would, make change, think anything, and would make	134.52 (6.97)	0.47
Store loyalty card/don't use store card			
SC16	Store loyalty card/don't use: card, store, loyalty, loyalty card, store card, health, store loyalty, link, card health, don't use, use loyalty, why store, no need, card would, and use store	147.65 (7.65)	0.34
"Big Brother" and privacy invasion			
SC8	Big brother/nanny state: feel, privacy, big brother, brother, nothing, invasion, invasion privacy, regard, don't feel, thing, feel like, state, choice, watch, and nanny	80.10 (4.15)	0.16
SC1	Privacy and cold calling: concern, privacy, concern about, confidentiality, would concern, email, require, call, future, bombard, advertise, market, guarantee, about privacy, and would require	73.15 (3.79)	0.13
Personal information sharing and access by others			
SC2	Concerns about linkage and insurance: information, give, idea, insurance, don't like, health, like idea, good, information could, wrong, reassurance, health insurance, hand, affect, and know about	80.87 (4.19)	0.05
SC19	Data access and others: information, private, personal, access, company, stored, detail, hold, information stored, sell, people, safe, personal information, personal health, and party	103.45 (5.36)	0.04
SC3	Don't want to share personal information: share, information, information share, personal, share information, personal information, don't want, nothing, detail, want information, nothing don't, want share, wouldn't want, don't like, and not share	100.36 (5.20)	0.12
Data inaccuracy			
SC7	Data inaccuracy and bias: buy, shop, make, purchase, family, people, food, eat, lifestyle, relate, supermarket, diet, healthy, product, and good	117.92 (6.11)	0.09
Data security and protection			
SC15	Don't trust organizations with data: data, trust, don't trust, NHS[c], share, organization, system, personal data, data share, hack, personal, guarantee, secure, not trust, and safety	95.34 (4.94)	0.09
SC12	Data protection: data, data protection, protection, data would, data use, access, issue, how data, health data, would use, health, wouldn't want, data link, secure, and link	81.06 (4.20)	0.14
SC9	Data security: security, data, data security, breach, worry, assurance, worry about, security data, data breach, about security, information security, increase, improve, risk, and security information	82.41 (4.27)	0.06
SC5	Guaranteed data safety/security: secure, 100, 100 percent, percent, convince, stored, safely, safe, stored safely, prefer, separate, if could, NHS, control, and how secure	74.88 (3.88)	0.37
Understanding research purpose and process			
SC20	Don't understand benefit: understand, benefit, don't understand, why would, would need, understand why, necessary, link, purpose, need understand, understand benefit, understand purpose, why need, would necessary, and need link	82.41 (4.27)	0.17
SC6	Require demonstratable benefits: benefit, not sure, benefit would, link, explanation, see benefit, explain, appropriate, explanation why, would benefit, explain benefit, if could, sure would, care, and sure why	69.48 (3.60)	0.03

XSL•FO
RenderX

Themes and SC[a] topic number	Topic name: top 15 terms selected by highest probability of the term given the topic	Prevalence (estimated survey responses; n=1930)[b], n (%)	Coherence
SC4[d]	Require reassurance: research, depend, data, purpose, specific, would depend, study, team, anonymize, happy, access, research team, if data, contact, and not use	92.06 (4.77)	0.07
SC18[d]	Require more information: information, more information, would use, information would, would need, would want, need know, information use, know why, want know, need more, would like, why would, information about, and how would	146.68 (7.60)	0.10
Health records shouldn't be linked			
SC17	Health record should not be linked: health, record, health record, link, link health, nothing, private, confidential, don't know, would link, why would, health care, nothing health, record would, and care	110.20 (5.71)	0.26
SC11[d]	Shopping habits and health shouldn't be linked: shop, health, habit, shop habit, interest, health care, link, wouldn't want, professional, shop health, condition, business, supermarket, health professional, and commercial	73.15 (3.79)	0.10
Don't understand reason/relevance of data linkage			
SC13	Unsure of reason for linkage: not sure, relevant, medical, would need, unsure, sure why, record, medical record, don't see, why would, relevant health, not relevant, sure would, medical information, and sure how	92.45 (4.79)	0.04
SC10	Don't see reason for linkage: reason, don't see, link, relevance, can't see, reason why, nothing, see relevance, point, see why, 2, see reason, should link, connection, and good	91.87 (4.76)	0.08

[a]SC: store card.

[b]Total prevalence does not sum exactly to 100%, and the total survey response counts do not sum exactly to N because of rounding.

[c]NHS: National Health Service.

[d]Topic has been regrouped to the category most thematically aligned with its contents from the category hierarchical clustering indicated as semantically similar.

XSL•FO
RenderX

Table 4. Topics created by latent Dirichlet allocation modeling of the LifeInfo health and fitness app question.

Themes and HA[a] topic number	Topic name: top 15 terms selected by highest probability of the term given the topic	Prevalence (estimated survey responses; n=1206)[b], n (%)	Coherence
Nothing would change mind			
HA17	Nothing would change mind: change, mind, change mind, nothing, would change, don't think, nothing would, think anything, make, wouldn't change, anything would, make change, would make, future, and think would	75.86 (6.29)	0.46
Apps/websites/wearable devices and don't use			
HA16	Device and don't use: device, use device, not use, wear, don't use, device not, wearable, collect, future, wearable device, data collect, wear device, app, data, and device future	59.09 (4.90)	0.10
HA11	App/website and don't use: app, fitness, use app, don't use, device, fitness app, lifestyle, app not, user, website, applicable, health app, not applicable, use fitness, and device app	53.67 (4.45)	0.13
"Big Brother" and privacy invasion			
HA3	"Big Brother:" feel, good, idea, life, big brother, brother, feel like, make, don't like, watch, like idea, bit, exercise, control, and NHS[c]	55.8 (4.62)	0.06
HA2	Privacy invasion and safety/security of data: privacy, safe, store, securely, invasion, not safe, store securely, invasion privacy, feel, hacker, issue, nothing, code, data store, and partly	53.43 (4.43)	0.05
Personal information sharing			
HA20	Information would be shared: information, share, information share, information wouldn't, store, know information, share information, wouldn't share, information store, sure information, worry, will share, health information, would share, and information will	65.12 (5.40)	0.04
HA14	Information is personal: personal, information, private, access, personal information, don't want, personal use, private information, information personal, reason, people, access information, point, long, and personal detail	61.75 (5.12)	0.08
Who has access to these data?			
HA15	Data access by insurance/private companies: company, insurance, lifestyle, relevant, make, insurance company, health, monitor, fitbit, will not, not relevant, point, interest, unsure, and wear	53.79 (4.46)	0.12
HA7	Health records and linkage: record, health, health record, access, link, doctor, don't want, information, access health, food, buy, not want, put, people, and link health	63.44 (5.26)	0.20
Data inaccuracy			
HA5	Inaccurate data and partial use: accurate, phone, not accurate, step, activity, app, hold, give, run, exercise, don't think, record, count, picture, and walk	52.58 (4.36)	0.11
Data security and protection			
HA19	Not sure and security: not sure, secure, sure how, sure would, would secure, sure about, secure would, convince, illegible, situation, how secure, sure anything, sure if, would convince, and if would	53.79 (4.46)	0.07
HA18	Don't trust data security: trust, secure, nothing, don't trust, computer, hack, fully, website, internet, not trust, nothing don't, nothing secure, wouldn't trust, information, and trust information	60.66 (5.03)	0.08
HA10[d]	Data protection against sharing: data, share, protection, data could, data protection, not share, data not, if data, share data, breach, system, sure data, thing, NHS, and bad	59.21 (4.91)	0.14
HA12[d]	Requires data security: security, data, concern, concern about, assurance, data security, safety, internet, information, about security, security information, use data, security would, assure, and matter	66.33 (5.50)	0.06
Understand research purpose and process			
HA6	Depends on assurances and purpose: depend, would depend, 100, 100 percent, percent, guarantee, depend how, depend information, depend use, depend why, give, depend if, depend purpose, percent guarantee, and how use	59.94 (4.97)	0.32

XSL•FO
RenderX

Themes and HA[a] topic number	Topic name: top 15 terms selected by highest probability of the term given the topic	Prevalence (estimated survey responses; n=1206)[b], n (%)	Coherence
HA4	Consent to specific research: research, happy, specific, permission, study, purpose, condition, time, project, if know, only if, researcher, advance, consent, and research project	58.37 (4.84)	0.08
HA9[d]	Understanding how and why data would be used: benefit, data, would use, understand, would need, don't know, data would, how would, want know, clear, would want, purpose, give, understand why, and benefit would	60 (4.98)	0.07
HA8[d]	Requires more information: information, would need, more information, detail, need more, need know, information would, would want, know more, why would, information use, anonymous, more about, more detail, and will use	72.84 (6.04)	0.11
HA1[d]	Benefits to health: health, benefit, researcher, link, professional, care, interest, health professional, health researcher, if health, health care, information, individual, would benefit, and don't see	66.21 (5.49)	0.06
Same answer as question 4 (store loyalty card question)			
HA13	Same answer as Q4: q4, answer, 4, question, previous, question 4, answer q4, see answer, response, previous answer, see previous, answer 4, answer question, response q4, and affect	54.15 (4.49)	0.12

[a]HA: health app.

[b]Total prevalence does not sum exactly to 100%, and the total survey response counts do not sum exactly to N because of rounding.

[c]NHS: National Health Service.

[d]Topic has been regrouped to the category most thematically aligned with its contents from the category hierarchical clustering indicated as semantically similar.

Making Recommendations

The recommended actions that researchers can take to address the key barriers to data sharing and linkage identified through LDA modeling are presented. These recommendations are based on the synthesis of participant suggestions and expertise regarding wider research on data sharing.

Results

Data Collection

The *LifeInfo Survey* recruited 7101 participants. The primary results of this study are reported elsewhere [28]. In brief, of those who reported using the services, 51.50% (2521/4895) responded favorably to sharing their loyalty card data for linkage to health records and 70.80% (1717/2425) responded favorably to sharing data from health and fitness apps or wearable devices to link with their health records. For the store loyalty card question, 62.28% (1489/2391) of respondents who answered *no* to whether they would share their data for linkage provided a free-text response. Of those who answered *not sure*, 66.02% (814/1233) provided a response. For the health and fitness app question, 50.82% (839/1651) of respondents who answered *no* to whether they would share their data for linkage provided a free-text response. Of those who answered *not sure*, 56.8% (565/995) provided a response.

A number of respondents who had either answered *yes* or did not provide an answer to whether they would share their data provided free-text responses that were included in the analysis (n=35 for the store loyalty card question and n=127 for the health and fitness app question). In total, 2338 individuals provided a free-text response to the store loyalty card question and 1531 for the health and fitness app question. Preprocessing steps and removing single-word responses reduced the number of responses to 1930 for the store loyalty card question and 1206 for the health and fitness app question. Single-word responses were considered separately and are shown in Tables 1 and 2.

Health and Fitness Modeling

Topic Number Selection

For the store loyalty card question, perplexity scores for each model of candidate number of topics (k) indicated that the best model has a number of topics within the range of k=20 to k=60, as the perplexity scores plateau at their minimum value within this range (Figure 1). For the health and fitness app question, perplexity scores were minimized to within the topic number range of 20-30 topics (Figure 2). The average probabilistic coherence scores within these ranges of k varied by only very small amounts, indicating that the models had near-equivalent topic quality (Figures 3 and 4). To maximize human interpretation and for comparability between the 2 questions, parsimonious models of 20 topics were chosen for both the store loyalty card question and health and fitness app question.

XSL•FO
RenderX

Figure 1. Perplexity scores for the models—LifeInfo store loyalty card question, 10-fold cross-validation of topic modelling to establish the optimal number of topics for latent Dirichlet allocation.

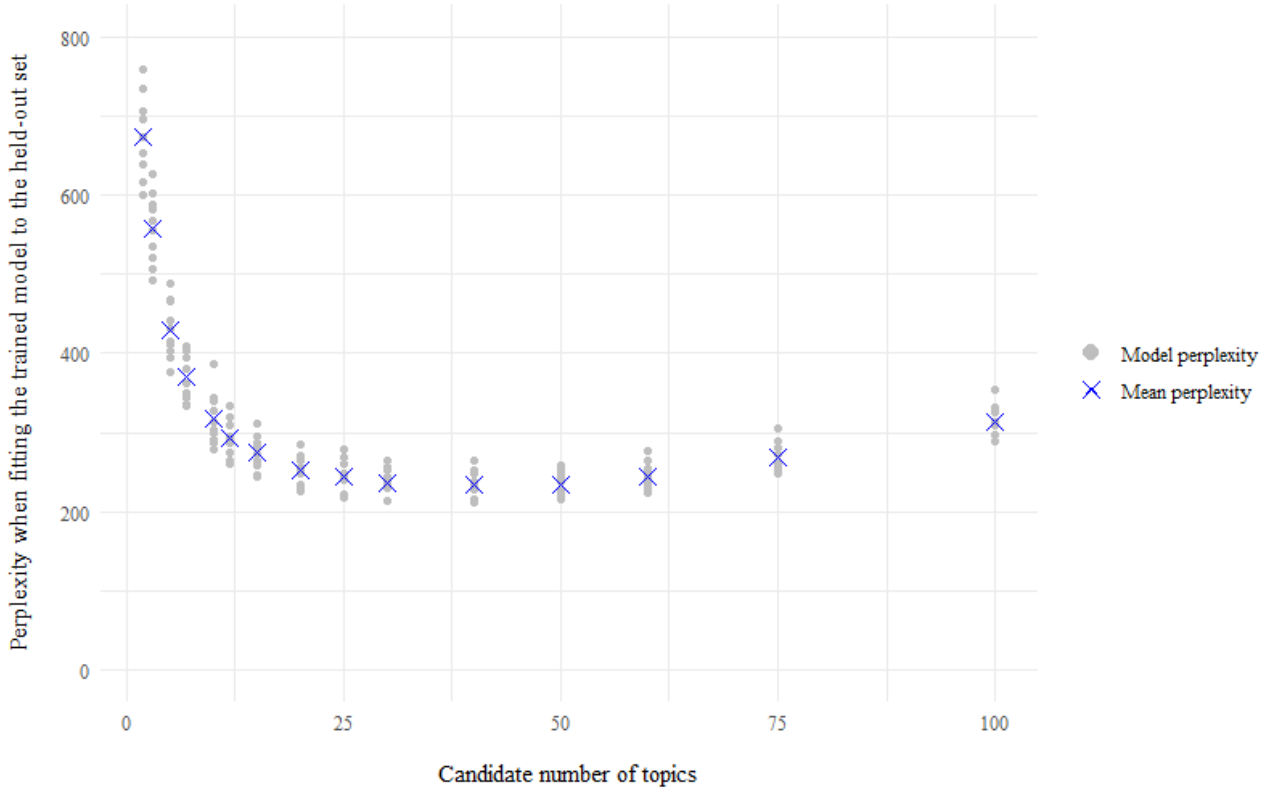
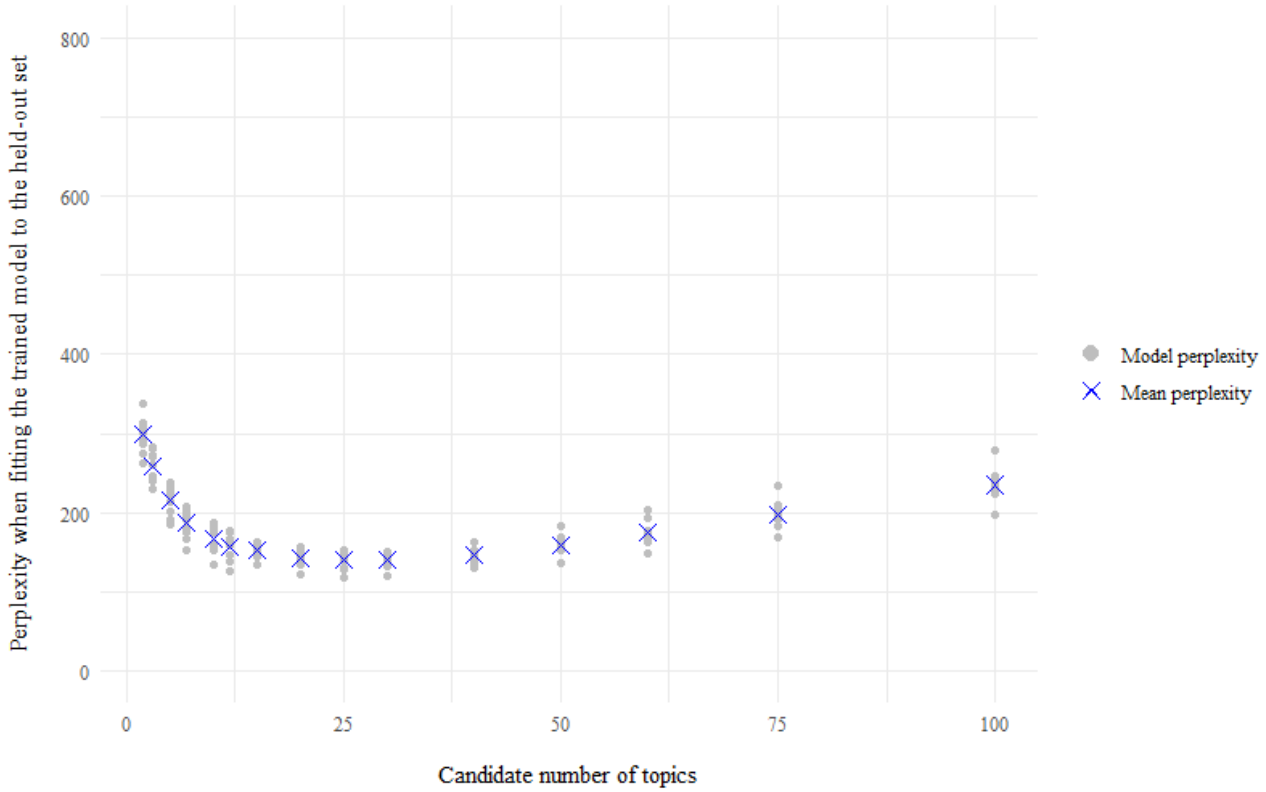


Figure 2. Perplexity scores for the models—LifeInfo health and fitness app question, 10-fold cross-validation of topic modelling to establish the optimal number of topics for latent Dirichlet allocation.



XSL•FO
RenderX

Figure 3. Average probabilistic coherence scores for the models—LifeInfo store loyalty card question, 10-fold cross-validation of topic modelling to establish the optimal number of topics for latent Dirichlet allocation.

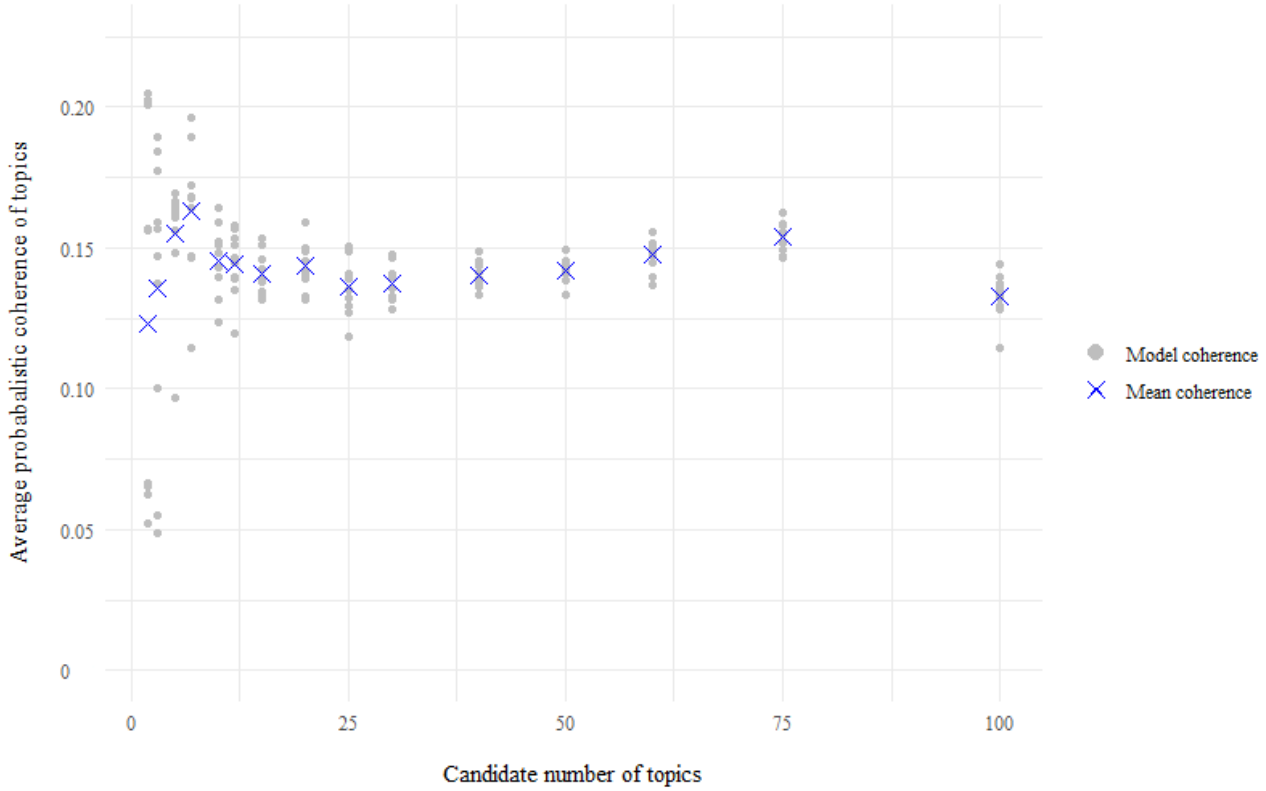
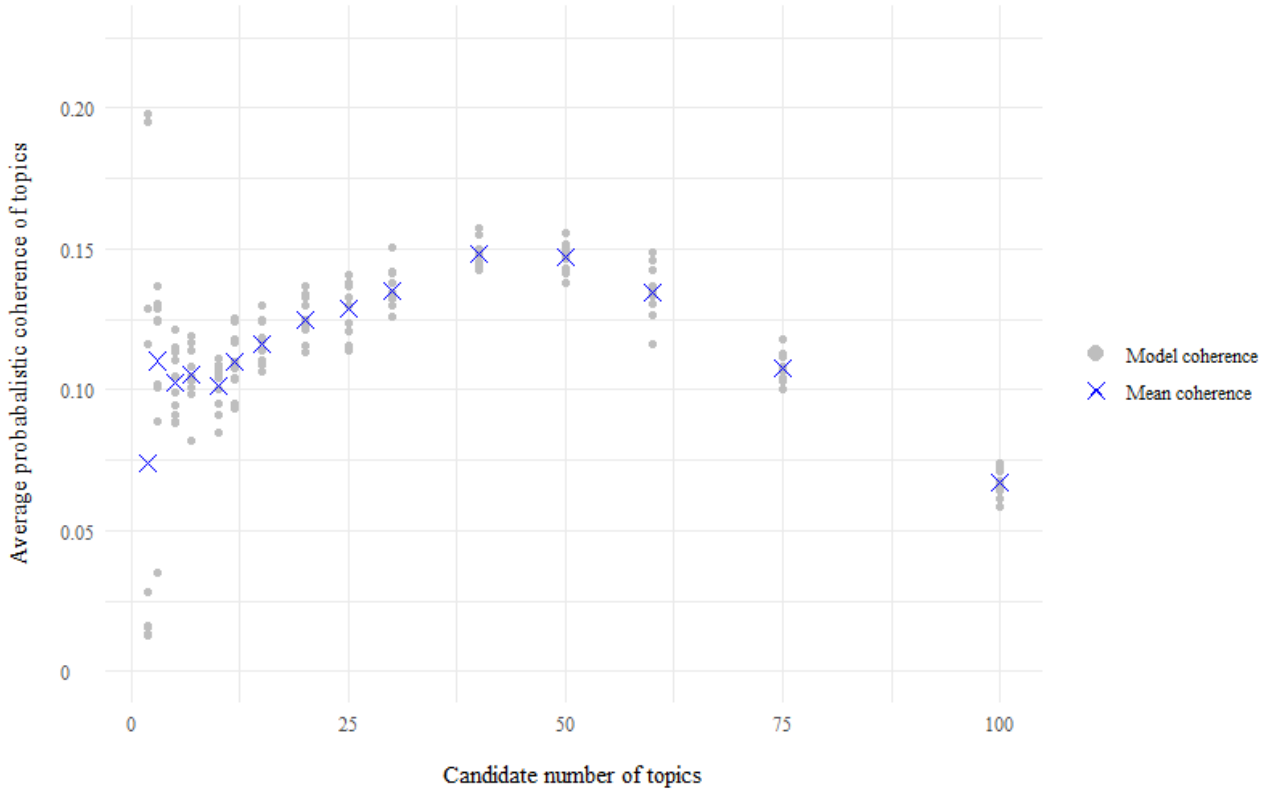


Figure 4. Average probabilistic coherence scores for the models—LifeInfo health and fitness app question, 10-fold cross-validation of topic modelling to establish the optimal number of topics for latent Dirichlet allocation.



XSL•FO
RenderX

LDA Topics

Table 3 shows the 20 topics created by the final LDA model of free-text responses to the store loyalty card question. The table includes the 15 top terms most representative of each topic, selected according to their φ values. As φ values are the probability of a term being used within a text given the topic, those terms with the highest φ values are representative of the topic content. The table includes 2 further topic measures: first, topic prevalence, which indicates how common each topic was within the survey responses (eg, prevalence 5.0 indicates 5% of survey responses fell into this topic), and second, probabilistic coherence, as an indicator of the semantic meaningfulness of each topic (this can range from 0 to 1, where higher numbers are more meaningful). Lower values of probabilistic coherence are owing to lower frequencies of topic term co-occurrences within texts and/or topic terms being highly frequent within the data set at large. Low coherence scores are associated with conceptually less defined topics [40]; however, they could also be caused by *fuzzy* topics with many different terms or semantic crossover within other topics. Table 4 shows the same information for the 20 topics created by LDA modeling of the health and fitness app question responses.

LDA modeling also assigns survey responses with probabilities for each topic, known as θ *values* A higher θ value for any given topic indicates that the response should more likely be categorized into that topic (Multimedia Appendix 5 shows example responses most associated with each topic). Some responses, especially those that are very short (consisting of only 1 substantive term), are given equal θ values across all topics, which can be regarded as *uncategorizable* (Multimedia Appendix 6). There are some differences in LDA topic frequency depending on demographic groups (eg, age and gender); further analysis demonstrating demographic patterns can be found in Multimedia Appendix 7. Topic stability is validated by comparing results across multiple LDA runs (Multimedia Appendix 8), indicating that the thematic categories reported are consistently produced in topics created through LDA.

Discussion

Principal Findings

Understanding attitudes toward using big lifestyle data for health research is important for the success of research initiatives interested in using these data in the future. More than half of our participants who generated big lifestyle data reported that they would be happy for these data to be linked to health records for future research. As only individuals who stated that they were unwilling or unsure about sharing their lifestyle data were prompted to respond to what would make them change their mind, these topics identified specific barriers to data sharing. Topic modeling on survey responses produced thematic topics that summarized latent themes of concern to potential data subjects. We believe that the intelligence generated will support researchers in addressing these issues in the future with the appropriate use of safeguards and consent procedures to generate a publicly acceptable study design.

It is also worth noting that, although LDA modeling aims to create distinct topics, individual responses may discuss multiple topics, and many of the topics identified were interconnected and complementary. In addition, topic quality varies and can be inferred by topic probabilistic coherence scores, which indicate how clearly defined each topic is semantically. For example, topic HA2 (Table 4) had a low value of probabilistic coherence (0.05) and primarily focused on the issue of privacy invasion; however, it also discussed the safety and security of data.

Barriers to Data Linkage

The topics uncovered by LDA modeling indicated that many of the same issues arise for both sharing store loyalty cards and health and fitness app data. Many topics can be matched across Tables 3 and 4 or highlight similar themes. Overall, key barriers to the use and linkage of store loyalty cards and physical activity data for health research included data safety and security, personal privacy, the need for further understanding about the research and study purpose, fear that data could get into the *wrong hands*, problems with data accuracy, and not understanding the reason for data linkage. These barriers can potentially be addressed by researchers with varying degrees of ease. However, for some respondents, nothing would make them share these data, whereas others did not use store loyalty cards or health and fitness apps. Many of these issues are common in the literature on health, consumer, and personal data sharing, and as such, these are expected findings; however, new concerns also arise specific to individual data linkages. Example responses most associated with each topic are used throughout this discussion and are labeled with their relevant topic. These responses were selected as those with the highest probabilities of being categorized into a given topic and can be viewed in Multimedia Appendix 5.

Nothing Would Change My Mind

Among those who responded negatively to sharing their lifestyle data for health research, a large proportion would be unwilling to change their mind. This is indicated first by the number of texts that LDA modeling categorized as the topic *nothing would change mind*, which refers to SC14 and HA17, constituting approximately 6.94% (134/1930) and 6.30% (76/1206) of the analyzed responses, respectively, and second, by the single-word answers excluded from the analysis, which were mainly the words *nothing*, *no*, and *none* for both the store loyalty card question and the health and fitness app question (Tables 1 and 2). These 3 single-word answers combined account for 90.3% (361/396) of the single-word responses about store loyalty cards and 85.8% (265/309) of single-word responses about health and fitness apps.

"Don't Use Services"

Others mentioned that they did not use these services and thus would not have data to share. This was most clearly identified within the topics related to health and fitness apps (HA16 and HA11) and was evident for store loyalty cards (SC16). For example:

> *I'm not going to use a wearable device in the future.* [HA16]

XSL•FO
RenderX

Currently don't use store loyalty cards. [SC16]

For this group, only greater participation in big lifestyle data production would allow them to share their data.

"Big Brother" and "Privacy Invasion"

Some respondents reported feeling that the proposed data linkage was a *big brother* and an invasion of privacy (SC8, SC1, HA2, and HA3). However, for the fitness app question, these topics were less clearly defined by LDA modeling, reflected in their low probabilistic coherence scores for these topics (0.05 and 0.06). Answers such as "not interested to a 'big brother is watching' on all aspects of my life" (SC8) and "[I] wouldn't want to feel every area of my life is out of my control and being watched by an institution that already makes me feel like I have no autonomy" (HA3) indicate a dislike and feeling of *surveillance* through the data. Actions by researchers to address these feelings are limited as they are related to a broader distrust of big data; however, greater transparency as to how data are being used and assurance of anonymity may convince some users. Within the responses about store loyalty cards, SC1 specifically identified privacy concerns related to unwanted emails, phone calls, and text that can be addressed through data protection and security actions.

Personal Data and Linkage

Many responses focused on concerns about sharing these data, which respondents perceived as highly personal information (HA20, HA14, SC19, and SC3). References to data being personal and confidential appeared across topic categories, particularly in the context of not wanting to store loyalty card data and health records linked, data protection and security, and concerns about who is able to access these data. For example:

> *Nothing, this information is for my personal use. Access to my private devices can lead to security risks.* [HA14]

The degree to which an individual believes that their data are personal and sensitive influences their willingness to share data and with whom. Medical information is personal, particularly for individuals with complex health conditions. Health and fitness app data and store loyalty card data are personal in different ways. Fitness apps are primarily used for personal monitoring, meaning data are not created with the idea that they might be shared with other actors, whereas for store loyalty card data, individuals exchange information about purchase history with shops in exchange for discounts and points. However, in the case of transaction data, Skatova et al [13] found that people regarded the graduality of transactional data as personal.

"Who Has Access to Private Data?"

Health data are regarded as particularly sensitive and confidential [26], related to the theme of *private information*, and many responses mentioned worries that data would be accessed by other actors without their permission (SC2, HA15, and HA7). For example, "if the information was available only to the research team, and not to others, e.g. insurance companies, mortgage companies, even the medical team" (HA15).

Private companies, third parties, and health insurance companies were frequently mentioned by respondents as actors they feared would gain access to their data. This is expected, as research has found that these institutions are trusted least by the public to use data appropriately [24,27]. Many respondents believed these companies would use their data for profit, to increase the cost of premiums, or to deny treatment altogether. However, respondents also mentioned concerns that health care professionals would be able to access their lifestyle data. For example:

> *It's my information for me. If I want my doc to know it I'll tell him/her or put it on my health record myself.* [HA7]

This finding is less expected, as research finds that health care providers are one of the most trusted actors for data sharing [24,27]. This indicates that there is no straightforward relationship between an individual's willingness to share their data and their trust in the actors involved. A further barrier for data linkage is illuminated in that data are often created for specific purposes, and alternative uses of data outside of this domain can create suspicion.

There was also a common misunderstanding that sharing lifestyle data for research would allow all involved actors to access these data. This was reflected in the previous example and others. For example, one respondent asked, "Why would I want Tesco knowing my health records?" (SC17).

Similar findings were reported by Skatova et al [13] in their research on transaction data sharing. This indicates that one easily achievable action that could influence people to share their data would be to explicitly state that data would not be available to anyone but researchers.

Closely related to data access by others was a reported general belief that health records should not be linked with other data and should be accessed only by health professionals (SC17 and SC11). This again reflects the belief that data should only be used for its designated purpose. For example, one respondent felt that "Health records should be kept in health care" (SC17).

Data Accuracy

Data accuracy was another identified barrier. Many respondents indicated that their purchase history or fitness tracking was only partial, creating misleading data about lifestyle behaviors (SC7 and HA5). Indeed, missingness and data integrity have been identified as a challenge for research using big lifestyle data and a concern for participants [10,26]. Nevertheless, good results have been found by comparing or combining these data with more traditional collection forms, for example, modeling individual consumption from household-level data [10,46].

When looking at responses within this topic, most participants reported concerns that their data would make them appear *less* healthy than their real activity. For example:

> *I buy all my fruit and veg at a farm shop...my data would provide misleading associations.* [SC7]
>
> *[my mobile data] shows a terrible step count, but that's because I don't hold my phone while playing netball, long walks etc.* [HA5]

XSL•FO
RenderX

Implicitly, respondents worried that they would be judged unfavorably on their lifestyle behaviors. Researchers could address these concerns by making explicit in the study protocol that (1) data are not expected to be complete, (2) detailed actions they will take to accommodate for this with modeling techniques or additional surveying, and (3) all data would be made unidentifiable foreclosing the possibility of judgment.

Data Protection and Security

Topics that focused on data protection and security formulated a large proportion of the responses (approximately 17.31% (334/1930) of responses for the store loyalty card question and 19.90% (240/1206) for the health and fitness app question). Changes frequently mentioned were assurances that these data would be stored completely safely and securely and that they would be protected from hacks and data breaches. These are common concerns for data sharing, especially with data that are regarded particularly private or sensitive, such as health data [26]. Although these assurances were given as a condition for data sharing within the question wording (Multimedia Appendix 1), many responses highlighted their importance or were not convinced that this would happen. Action would, therefore, need to be taken to give participants greater confidence, which could be achieved through transparency in the research process and providing details of how these data will be protected.

Understanding More About the Research Purpose, Process, and Benefit

Across both store loyalty cards and health and fitness app data sharing, respondents mentioned that understanding the research better and being given more control over their participation would influence them to change their mind. This includes being provided with more information (SC18 and HA8), giving permission only for specific research projects (SC4 and HA4), and a greater understanding of the reason or benefits of research (SC20, SC4, SC6, HA9, and HA1). These findings are supported by the research of Skatova et al [13], who found that support for data sharing is contingent on its context and purpose, highlighting the importance of well-informed participants. Given that the *LifeInfo Survey* aimed to assess attitudes toward data sharing in the future, rather than requesting participants to consent to data sharing at this time, the participant information sheet (Multimedia Appendix 9) needed to be suitably broad, which would not be the case when recruiting participants to an actual data linkage study.

Researchers have found that a strong social motive, such as improvements to health or treatment, in addition to clearly defining the purpose of research, are key motivating factors for personal data donation [47], whereas using health data for insurance, marketing purposes, or commercial exploitation is unacceptable to the public [23]. This was reflected in the unstructured answers from the *LifeInfo Survey*. A respondent answered that they would be supportive "if the data fed into important public health or similar research and was not used to further commercial gains by these giants of commerce" (HA4).

Again, although LDA separates topics, they are connected, and the involvement of private companies creates concerns about whether these entities will be able to access health data for profit once linked.

Control and consent for health data being used for research have been found to be key for public acceptance [25]. A respondent answered that the data "has to be linked to a condition or specific research project with additional approval provided in advance" (HA2).

This is something that can be adapted into the research process, allowing participants to consent to or deny the use of their data, given the specifications of the study.

In addition to a desire for more understanding, another thematic category was identified of individuals who did not understand the reason or relevance of data linkage (SC13 and SC10). For example, one respondent said, "Do not see any reason why they should be linked" (SC10).

These topics were only produced when modeling the store loyalty card question, although topic HA5 also encompassed responses that stated data are *not relevant*. This perhaps indicates that respondents found the link between purchasing and health to be less relevant than fitness tracking and health.

Strengths and Limitations

The application of LDA modeling has clear strengths, enabling semiautomated analysis of large text corpora to readily identify barriers for data sharing. Nonetheless, some limitations present themselves; the topics identified through LDA modeling may hide rarer topics that do not have highly frequent mentions or a homogenous lexicon; for example, some *LifeInfo Survey* responses mention financial compensation (eg, vouchers), but this is not identified as a topic. In addition, texts that are linked by similar terms but are thematically different are sometimes grouped by LDA modeling; for example, HA6 included responses that require *100 percent* reassurance of certain criteria. However, these criteria span several different issues. These more granular findings may be better identified by human qualitative coding; however, this comes with its own limitations, especially for large data sets.

The *LifeInfo Survey* sample size was large, thereby facilitating the identification of important topics; however, the size of this data set was smaller than those often used for topic modeling, and responses were relatively short. As previously mentioned, those texts that were extremely short were uncategorizable by the model, which is a limitation of this methodology and data. Similarly, the survey was designed to elicit responses only from those unwilling or unsure about sharing their data. This provided benefits as it focused on the scope of topics to identify key barriers; yet, it would be insightful to obtain the opinions of those more supportive of lifestyle data sharing initiatives (which was 52.30% (2521/4820) of loyalty card holders and 70.80% (1717/2425) of health and fitness app users in our study), which should be considered in future studies. In examples where positive and negative responses are captured, it would be useful to explore a sentiment analysis approach to text mining; however, this was not relevant for our study.

Research has found LDA model results to be sensitive to model hyperparameters [44], and it is possible to use methods that

XSL•FO
RenderX

optimize LDA across different α and β values. These methods were not applied in this study as trialing them increased the computational intensity of the analysis and did not provide better solutions.

Due to resource limitations, the *LifeInfo Survey* questionnaires were only available in English. This means that we are unlikely to have reached the <2% of the population who are not able to speak English. However, the *LifeInfo Survey* is overrepresented in the traditionally hard-to-reach, most deprived communities and Asian and other ethnicities [28].

Recommended Actions

Several actions can be taken by researchers to directly address the key barriers to data sharing identified through LDA modeling, a summary of which is detailed in Textbox 1.

Textbox 1. Summary of recommendations to improve support for the linkage of novel consumer and lifestyle data with health records for research purposes.

Motivation

- Provide detailed and specific information about the study purpose and benefit.

Control and consent

- Provide detailed information about research and specific opt-in mechanisms to give participants more control.

Access by others

- Provide an explicit statement that data linkage does not give all parties access to linked data. Lifestyle data will not be shared with health services, and health records will not be shared with supermarkets or technology companies.

Third-party access

- Provide reassurances that data will not be shared with third parties, such as health insurers.

Inaccurate data

- Provide acknowledgment within the study specification that data might be partial and outline mechanisms for how this will be addressed, such as data quality checks, modeling techniques, and supplementary data collection.

Non and/or infrequent use of services

- Increase participation in novel data collection and more complete use.

Data security and protection

- Put in place stringent precautions to keep data protected from hacks or data breaches.

Personal data

- Provide assurances that data will be made anonymous and nonidentifiable.

“Big Brother” and privacy invasion

- Widely used good practice and exemplar studies, which provide a clear benefit to public health, and excellent data security could help increase trust in data sharing initiatives.

Findings and recommended actions incorporate some areas we hypothesized would emerge, for example, data security and protection, but are far more comprehensive and nuanced than the existing literature reflects.

Future Work

There is extensive scope to use LDA in future research, where free-text responses are collected in large surveys. LDA allows the detection of topics in free-text responses to be generated at a scale that is not feasible in more traditional qualitative thematic methods. Other text mining approaches, such as sentiment analysis, may be applicable where identification of positive and negative responses is important.

More work is needed to be able to further unpick concerns within topics such as *who has access to data*, where this could be context-dependent, for example, patients with a complex medical history who are worried about being judged on their lifestyle behaviors by their clinical team, and *big brother*, where it is clear that greater transparency is required regarding data being used for research purposes.

Future research could explore the utility of encouraging patients or research participants to use store loyalty cards and health and fitness apps or wearables as part of their personalized care, extending research that is already being done. This is of particular significance for those relating to the *don’t use services* topic.

There has been significant interest in the use of supermarket loyalty cards and health and fitness app data in health research, in addition to the greater availability of these data in recent years. This provides exciting opportunities to gain new insights into lifestyle risk factors for diseases through individual-linked

XSL•FO
RenderX

data. The growth of this research requires that the common concerns of participants regarding ethics, data security, research aims, and personal privacy, among others, are understood so that they can be addressed by future projects. Researchers may use the findings and recommended actions shared in this paper so that greater trust can be built in practices of data linkage for health research.

Conclusions

Analysis of the *LifeInfo Survey* responses with topic modeling techniques revealed key barriers that prevent people from willingly sharing their novel lifestyle data for health research. This large-scale public consultation provides actionable recommendations that will allow researchers using big lifestyle data to adapt their study design and provide safeguards based on expressed concerns important to the general public that are specific to novel lifestyle data and health record linkage.

Acknowledgments

The authors would like to thank all the LifeInfo study participants. The authors would also like to thank the research teams members at Leeds Teaching Hospitals National Health Service Trust, Kinga Dwornik, Patric Devitt, Linda Bamford, Ian Thompson, Amanda Friend, Linda Roberts, Richard Evans, and Laura Kelly; Leeds Institute for Data Analytics colleagues, Kimberley Wright, Hayley Irving, Chris Carrigan, and Maria Galazoula; Low Moor Medical Practice colleagues, Mark Cade and Maureen Rowland; Leeds City Council colleague, Thomas Woolley; and the UseMyData group. The LifeInfo study has been supported by the National Institute for Health Research Clinical Research Network through portfolio adoption, the Consumer Data Research Centre grant ref:ES/L011891/1, the Medical Bioinformatics Centre ref: MR/L01629X/ and the School of Medicine, University of Leeds, through their Academic Development Fund.

Conflicts of Interest

MM is an inventor and shareholder at Dietary Assessment Ltd.

Multimedia Appendix 1
LifeInfo questionnaire.
[PDF File (Adobe PDF File), 217 KB - jmir_v23i5e24236_app1.pdf]

Multimedia Appendix 2
Supplementary information regarding data cleaning and processing steps.
[DOCX File , 14 KB - jmir_v23i5e24236_app2.docx]

Multimedia Appendix 3
Question response lengths before and after data processing.
[DOCX File , 28 KB - jmir_v23i5e24236_app3.docx]

Multimedia Appendix 4
Dendrograms of hierarchical clustering of topics produced from latent Dirichlet allocation models.
[DOCX File , 91 KB - jmir_v23i5e24236_app4.docx]

Multimedia Appendix 5
Tables showing the 20 topics produced by latent Dirichlet allocation modeling and the top 5 words per topic according to the response θ value.
[DOCX File , 35 KB - jmir_v23i5e24236_app5.docx]

Multimedia Appendix 6
Summary of texts that are not categorizable into any specific topic by the latent Dirichlet allocation model.
[DOCX File , 13 KB - jmir_v23i5e24236_app6.docx]

Multimedia Appendix 7
Graphs showing mean topic prevalence and standard error bars broken down by demographic groups.
[DOCX File , 74 KB - jmir_v23i5e24236_app7.docx]

Multimedia Appendix 8
Coded tables of multiple latent Dirichlet allocation outputs to test topic stability.

XSL•FO
RenderX

[DOCX File , 43 KB - jmir_v23i5e24236_app8.docx]

Multimedia Appendix 9
Participant information sheet.
[PDF File (Adobe PDF File), 271 KB - jmir_v23i5e24236_app9.pdf]

References

1. Afshin A, Sur P, Fay K, 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;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
2. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9]
3. Physical activity: applying all our health. Public Health England. 2019. URL: https://www.gov.uk/government/publications/physical-activity-applying-all-our-health/physical-activity-applying-all-our-health [accessed 2020-12-20]
4. The food system in data - Part One. National Food Strategy. URL: https://www.nationalfoodstrategy.org/partone/ [accessed 2020-12-20]
5. Timmins KA, Green MA, Radley D, Morris MA, Pearce J. How has big data contributed to obesity research? A review of the literature. Int J Obes (Lond) 2018 Dec 18;42(12):1951-1962 [FREE Full text] [doi: 10.1038/s41366-018-0153-7] [Medline: 30022056]
6. Morris MA, Wilkins E, Timmins KA, Bryant M, Birkin M, Griffiths C. Can big data solve a big problem? Reporting the obesity data landscape in line with the Foresight obesity system map. Int J Obes (Lond) 2018 Dec 21;42(12):1963-1976 [FREE Full text] [doi: 10.1038/s41366-018-0184-0] [Medline: 30242238]
7. Green MA, Watson AW, Brunstrom JM, Corfe BM, Johnstone AM, Williams EA, et al. Comparing supermarket loyalty card data with traditional diet survey data for understanding how protein is purchased and consumed in older adults for the UK, 2014-16. Nutr J 2020 Aug 13;19(83):1-10 [FREE Full text] [doi: 10.1186/s12937-020-00602-3] [Medline: 32791968]
8. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun 01;37(3):197-206 [FREE Full text] [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]
9. Bidargaddi N, Musiat P, Makinen V, Ermes M, Schrader G, Licinio J. Digital footprints: facilitating large-scale environmental psychiatric research in naturalistic settings through data from everyday technologies. Mol Psychiatry 2017 Dec;22(2):164-169 [FREE Full text] [doi: 10.1038/mp.2016.224] [Medline: 27922603]
10. Nevalainen J, Erkkola M, Saarijärvi H, Näppilä T, Fogelholm M. Large-scale loyalty card data in health research. Digit Health 2018 Nov 29;4:- [FREE Full text] [doi: 10.1177/2055207618816898] [Medline: 30546912]
11. Aiello LM, Schifanella R, Quercia D, Del Prete L. Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci 2019 Apr 30;8(14):1-22. [doi: 10.1140/epjds/s13688-019-0191-y]
12. About the CDRC. Consumer Data Research Centre. URL: https://www.cdrc.ac.uk/about/ [accessed 2020-12-06]
13. Skatova A, Shiells K, Boyd A. Attitudes towards transactional data donation and linkage in a longitudinal population study: evidence from the Avon Longitudinal Study of Parents and Children. Wellcome Open Res 2019 Dec 3;4:192. [doi: 10.12688/wellcomeopenres.15557.1]
14. Mamiya H, Moodie EE, Buckeridge DL. A novel application of point-of-sales grocery transaction data to enhance community nutrition monitoring. AMIA Annu Symp Proc 2017;2017:1253-1261 [FREE Full text] [Medline: 29854194]
15. Clark S, Shute B, Jenneson V, Rains T, Morris M. Compliance with the Eatwell guide: a case study using supermarket transaction records in Yorkshire and the Humber. Proc Nut. Soc 2020 Jun 10;79(OCE2):e665. [doi: 10.1017/S002966512000614X]
16. Jenneson V, Shute B, Greenwood D, Clarke G, Clark S, Rains T, et al. Variation in fruit and vegetable purchasing patterns in Leeds: using novel loyalty card transaction data. Proc Nutr Soc 2020 Jun 10;79(OCE2):e670. [doi: 10.1017/S0029665120006199]
17. Gay V, Leijdekkers P. Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability. J Med Internet Res 2015 Nov 18;17(11):e260 [FREE Full text] [doi: 10.2196/jmir.5094] [Medline: 26581920]
18. Brinton JE, Keating MD, Ortiz AM, Evenson KR, Furberg RD. Establishing linkages between distributed survey responses and consumer wearable device datasets: a pilot protocol. JMIR Res Protoc 2017 Apr 27;6(4):e66 [FREE Full text] [doi: 10.2196/resprot.6513] [Medline: 28450274]
19. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. npj Digit. Med 2019 Jun 3;2(1):1-12. [doi: 10.1038/s41746-019-0121-1]
20. Minerva. Problems with care.data and other stories. Br Med J 2015;351:-. [doi: 10.1136/bmj.h4613]
21. Sweeney L. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University. 2000. URL: http://dataprivacylab.org/projects/identifiability/paper1.pdf [accessed 2021-03-04]

XSL•FO
RenderX

22. Ritchie F, Smith J. Confidentiality and linked data. arXiv. URL: https://gss.civilservice.gov.uk/guidances/quality/nsqr/privacy-and-data-confidentiality- [accessed 2021-03-04]
23. Patel R. Foundations of fairness: where next for NHS health data partnerships? Ada Lovelace Institute. 2020 Mar. URL: https://understandingpatientdata.org.uk/sites/default/files/2020-03/Foundations%20of%20Fairness%20-%20Summary%20and%20Analysis.pdf [accessed 2021-05-06]
24. ODI survey reveals British consumer attitudes to sharing personal data. Open Data Institute. 2018. URL: https://theodi.org/article/odi-survey-reveals-british-consumer-attitudes-to-sharing-personal-data/ [accessed 2020-05-07]
25. 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]
26. Summary report of qualitative research into public attitudes to personal data and linking personal data : summary report / Wellcome Trust. Wellcome Trust (London, England). 2013. URL: https://wellcomelibrary.org/item/b20997358#?c=0&m=0&s=0&cv=0 [accessed 2020-12-08]
27. Royal Society, Trust Deficiet -Lessons-for-Policymakers. Royal Statistical Society (RSS). URL: http://www.statslife.org.uk/news/1672 [accessed 2019-07-01]
28. Morris M. Public attitude to linking lifestyle data to health records for research: the LifeInfo study. Forthcoming. 2021. URL: https://osf.io/4bmpe/ [accessed 2021-05-10]
29. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural topic models for open-ended survey responses. Am J Pol Sci 2014 Mar 06;58(4):1064-1082. [doi: 10.1111/ajps.12103]
30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Machi Learn Res 3. 2003. URL: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf [accessed 2021-05-06]
31. Lund M, Williams K. Summary report on the 2013 Healthy Communities survey. Leeds City Council. 2013. URL: https://leedsobs.wpengine.com/wp-content/uploads/2018/03/Healthy-Communities-Survey-2013-Analysis-Report.pdf [accessed 2020-10-01]
32. Morris M. LifeInfo study. Open Science Framework. URL: https://osf.io/2rsnd/ [accessed 2020-09-14]
33. R: a language and environment for statistical computing. R Software. 2019. URL: https://www.r-project.org [accessed 2020-06-01]
34. Lau JH, Baldwin T, Newman D. On collocations and topic models. ACM Trans Speech Lang Process 2013 Jul;10(3):1-14. [doi: 10.1145/2483969.2483972]
35. Wallach H. Topic modeling: beyond bag-of-words. In: Proceedings of the 23rd international conference on Machine learning. 2006 Presented at: 23rd international conference on Machine learning; 2006; Carnegie Mellon University, Pittsburgh, Pennsylvania URL: http://dirichlet.net/pdf/wallach06topic.pdf
36. Nokel M, Loukachevitch N. A method of accounting bigrams in topic models. In: Proceedings of the 11th Workshop on Multiword Expressions. 2015 Presented at: 11th Workshop on Multiword Expressions; June 2015; Denver, Colorado. [doi: 10.3115/v1/w15-0901]
37. Greene D, O'Callaghan D, Cunningham P. How many topics? Stability analysis for topic models. In: Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer; 2014:498-513.
38. Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P. The author-topic model for authors and documents. arXiv.org. 2012. URL: http://arxiv.org/abs/1207.4169 [accessed 2020-07-13]
39. Chang J, Boyd-Graber J, Gerrish S, Wang C, Blei D. Reading tea leaves: how humans interpret topic models. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems. 2009 Presented at: 23rd Annual Conference on Neural Information Processing Systems; December 7-10, 2009; Vancouver, British Columbia, Canada URL: http://rexa.info
40. Mimno D, Wallach H, Talley E, Leenders M, Mccallum A. Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July, 2011; Edinburgh, Scotland, UK URL: https://www.aclweb.org/anthology/D11-1024
41. Jones TW. Topic modeling. 2019. URL: https://cran.r-project.org/web/packages/textmineR/vignettes/c_topic_modeling.html [accessed 2020-07-13]
42. Rosner F, Hinneburg A, Röder M, Nettling M, Both A. Evaluating topic coherence measures. Cornell University. 2014. URL: http://arxiv.org/abs/1403.6397 [accessed 2020-07-13]
43. 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 Presented at: Eighth ACM International Conference on Web Search and Data Mining; February, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
44. Wallach H, Mimno D, Mccallum A. Rethinking LDA: why priors matter. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems. 2009 Presented at: 22nd International Conference on Neural Information Processing Systems; December, 2019; Vancouver, B.C., Canada p. 1973-1981 URL: https://dl.acm.org/doi/abs/10.5555/2984093.2984314 [doi: 10.5555/2984093.2984314]
45. Upton G, Cook I. A dictionary of statistics. Oxford: Oxford University Press; 2014.

XSL•FO
RenderX

46. Becker W. Comparability of household and individual food consumption data--evidence from Sweden. Public Health Nutr 2001 Oct;4(5B):1177-1182. [Medline: 11924944]
47. Skatova A, Goulding J. Psychology of personal data donation. PLoS One 2019 Nov 20;14(11):e0224240 [FREE Full text] [doi: 10.1371/journal.pone.0224240] [Medline: 31747408]

Abbreviations

DTM: document term matrix
LDA: latent Dirichlet allocation

Edited by G Eysenbach; submitted 18.09.20; peer-reviewed by J Nevalainen, Z Ren, MDG Pimentel, I Shubina, N Buechler; comments to author 06.11.20; revised version received 27.01.21; accepted 12.04.21; published 17.05.21.

Please cite as:
Clarke H, Clark S, Birkin M, Iles-Smith H, Glaser A, Morris MA
Understanding Barriers to Novel Data Linkages: Topic Modeling of the Results of the LifeInfo Survey
J Med Internet Res 2021;23(5):e24236
URL: https://www.jmir.org/2021/5/e24236
doi:10.2196/24236
PMID:33998998

©Holly Clarke, Stephen Clark, Mark Birkin, Heather Iles-Smith, Adam Glaser, Michelle A Morris. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

HIV Information Acquisition and Use Among Young Black Men Who Have Sex With Men Who Use the Internet: Mixed Methods Study

Megan Threats[1,2], PhD, MSLIS; Keosha Bond[2,3], EdD, CHES, MPH

[1]School of Communication and Information, Rutgers University, New Brunswick, NJ, United States

[2]Yale School of Public Health, Center for Interdisciplinary Research on AIDS, New Haven, CT, United States

[3]Department of Community Health and Social Medicine, City University of New York School of Medicine, New York, NY, United States

Corresponding Author:
Megan Threats, PhD, MSLIS
School of Communication and Information
Rutgers University
4 Huntington Street
New Brunswick, NJ, 08901
United States
Phone: 1 848 932 7524
Email: megan.threats@rutgers.edu

Abstract

Background: HIV disproportionately affects young Black men who have sex with men (YBMSM) in the United States. eHealth holds potential for supporting linkage and engagement in HIV prevention and care and the delivery of HIV information to YBMSM.

Objective: This study aims to investigate HIV information acquisition and use among YBMSM who use the internet.

Methods: A web-based self-administered survey and semistructured interviews were conducted. The survey findings informed the development of the interview guide. Descriptive statistics were used to characterize the survey sample, and interview data were analyzed thematically using modified grounded theory methodologies.

Results: Among the internet sample (N=83), the average age was 29.2 (SD 3.5) years, 41% (n=34) of participants self-reported living with HIV, 43% (n=36) were HIV-negative, and 15% (n=13) were unsure of their HIV status. Most participants (n=79, 95%) acquired HIV information through the internet while using a mobile phone. Web-based HIV information was intentionally sought from consumer health information websites (n=31, 37%), government health information websites (n=25, 30%), and social media (n=14, 17%). Most men incidentally acquired HIV information via advertisements on social media sites and geospatial dating apps (n=54, 65%), posts on social media sites from their web-based social ties (n=44, 53%), and advertisements while browsing the internet (n=40, 48%). Although the internet is the top source of HIV information, health care providers were the most preferred (n=42, 50%) and trusted (n=80, 96%) source of HIV information. HIV information was used to facilitate the use of HIV prevention and care services. The qualitative sample included YBMSM across a range of ages and at different points of engagement in HIV prevention and care. Qualitative findings included the importance of the internet as a primary source of HIV information. The internet was used because of its ease of accessibility, because of its ability to maintain anonymity while searching for sensitive information, and to mitigate intersecting stigmas in health care settings. Participants used HIV information to assess their risk for HIV and AIDS, support their skill building for HIV prevention, inform patient–doctor communication, and learn about HIV prevention and treatment options. Men expressed concerns about their diminishing access to online spaces for HIV information exchange among YBMSM because of censorship policies on social media sites and the *stigmatizing* framing and tone of mass media HIV-prevention advertisements encountered while using the internet.

Conclusions: YBMSM in this sample had high utilization of eHealth for HIV information acquisition and use but diminished access to their preferred and most trusted source of HIV information: health care providers. Future eHealth-based HIV interventions culturally tailored for YBMSM should aim to reduce intersectional stigma at the point of care and support patient–provider communication. The findings demonstrate the need for community-informed, culturally tailored HIV messaging and online spaces for informational support exchange among YBMSM.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e22986) doi:10.2196/22986

KEYWORDS

HIV; health information behavior; eHealth; mHealth social media; consumer health informatics; mobile phones; sexual and gender minorities; African Americans; young adults; mixed methods

Introduction

Background

Young Black men who have sex with men (YBMSM) are disproportionately affected by HIV in the United States, accounting for more HIV diagnoses than men who have sex with men (MSM) of other ethnoracial groups [1,2]. Black MSM have a 50% chance of acquiring HIV in their lifetime, and if the current incidence rates persist, 40% of Black MSM will be diagnosed with HIV by the age of 30 years [3,4]. YBMSM experience disparities across every stage of the HIV prevention and care continuum [5-8]. Compared with MSM of other ethnicities, YBMSM are less likely to have knowledge of their HIV status and are more likely to be infected, yet unaware [1,6]. Biomedical advancements in HIV prevention and treatment, such as pre-exposure prophylaxis (PrEP), provide an opportunity to reduce HIV transmission and related disparities, but awareness and adoption of these advancements remain the lowest among YBMSM [9-14].

Related Literature

HIV-related information is a critical resource for people living with HIV (PLWH) and those at high risk for HIV. Assessing and managing the risk of acquiring or transmitting HIV requires the successful acquisition and application of accurate HIV information. As possessing HIV information does not automatically lead to health behavior change, individuals must be informed before they are motivated and equipped with the skills to perform HIV risk-reduction behaviors [15,16]. Improving our knowledge of how YBMSM acquire and use HIV-related information outside of planned interventions will aid in efforts to disseminate information and messages among this group, which is at an elevated risk for HIV.

There is a growing body of literature that examines health information behavior (eg, health information acquisition, sharing, management, and use) within the domain of HIV [17-20]. Much of the literature examining HIV-related health information behavior has been explored among samples consisting primarily of PLWH, heterosexual youth, and individuals living outside of the United States [21-27]. Consequently, such findings are not generalizable to YBMSM. For instance, because the HIV epidemic disproportionately affects YBMSM, heterosexual young adults often seek information about contraception (eg, pregnancy prevention) and less information about HIV [28-30]. Furthermore, previous research examining HIV information acquisition among Black MSM was conducted at a time when the internet was not as accessible via wireless connection and mobile devices [31].

There is an opportunity to leverage eHealth modalities (eg, the internet, mobile technologies, social media, video games, geospatial networking apps, and text messaging) to provide access to accurate HIV information and support its use among YBMSM [32-36]. Although there are several technology-enabled HIV interventions that provide access to credible HIV information and aim to improve HIV-related outcomes among YBMSM, most are specifically for YBMSM living with HIV [37-40]. Few eHealth or mobile health (mHealth) HIV interventions are inclusive of YBMSM of varying HIV status [41-46]. To inform the design of future eHealth and mHealth interventions that provide HIV-related information and encourage its application among YBMSM, we need a better understanding of how YBMSM currently use these technologies to access and interact with HIV information and services outside of planned interventions. This paper describes a part of a larger study on HIV-related information behaviors of YBMSM and factors affecting their engagement along the HIV prevention and care continuum [47]. This study aims to explore HIV information acquisition and use among YBMSM who use the internet.

Methods

Mixed Methods Study Design

A sequential, explanatory mixed methods design guided this study [48]. During the first phase of the study, a convenience sample of 83 YBMSM completed a web-based, self-administered survey. The survey data were then analyzed to help inform the development of the interview guide. In the second phase of the study, semistructured interviews were conducted among a subsample (n=22) of survey respondents to help elaborate on the survey findings and provide new insights. The survey and interview data were integrated or *mixed* during both the collection and analysis of data. The results from the quantitative and qualitative data analyses were triangulated during interpretation.

Participants and Recruitment

A convenience sample of 83 YBMSM was recruited from December 18, 2018, to May 28, 2019, through a variety of venues such as HIV or sexually transmitted infection (STI) clinics, AIDS Service Organizations, a convenience sample from the UNC Center for AIDS Research, social media sites, geospatial dating apps, bars or clubs, lesbian, gay, bisexual, trans, and queer (LGBTQ) organizations, historically Black colleges and universities, mobile instant messaging apps (GroupMe and WhatsApp), and public postings. The eligibility criteria included the following: self-identification as Black or African American; identification as gay, bisexual, same-gender loving, or a man who has had sex with other men; being aged between 18 and 34 years; residing in the state of North Carolina; willing and able to provide informed consent; and being able to read, write, and speak English. The study received ethical approval from the institutional review board of the University of North Carolina at Chapel Hill. A certificate of confidentiality was obtained from the National Institutes of Health for the study.

XSL•FO
RenderX

Part I: Internet Survey

The survey adapted questions from 3 previous surveys: the National Cancer Institute's Health Information National Trends Survey, the Centers for Disease Control and Prevention's National HIV Behavioral Surveillance Questionnaire, and other published survey items on HIV-related health information behavior [49,50]. The HIV-related health information behavior survey items selected for inclusion were based on theoretical models of information behavior, including an adapted version of the Wilson model of information behavior [51] and the Erdelez model of information encountering [52]. The survey was pilot tested on a community advisory board comprising YBMSM (data not included in the final analysis). The final survey contained 72 possible questions and took an average of 15.7 minutes to complete. In addition to sociodemographic information, the survey content included questions regarding technology ownership or use, HIV information behaviors, HIV or STI testing, HIV or STI treatment, and PrEP use. The web-based survey was administered using Qualtrics software (Qualtrics).

A total of 225 interested parties clicked on the Qualtrics link to read about the survey. Among these, 45.3% (102/225) consented to participate in the survey. Among these 102 participants, 11 (10.7%) were not eligible as they identified as female (5/102, 4.9%), 2 (1.9%) were aged above 34 years, and 4 (3.9%) reported a zip code outside of the state of North Carolina. Of the remaining 91 eligible participants, 83 (91%) completed the survey.

Descriptive statistics of the survey measures were conducted using Microsoft Excel (Microsoft Corporation) to summarize survey data, including sociodemographic characteristics and technology ownership or usage. In this paper, the reported HIV-related information behaviors included information seeking, incidental information acquisition, and information use. To assess HIV information seeking, 3 measures were used: decision to take action (eg, seeking motivators), selection of information source (eg, preference and trust), and sources of HIV information. To measure the frequency of incidental HIV information acquisition, participants were asked to respond how much times ("never," "a little," "some," or "a lot") in the past 2 years they had received HIV information in the following ways: "I learn unexpected things about HIV while browsing the Internet." Additional response options were reflected in the data. HIV information use was measured by asking participants to identify the ways in which they applied the HIV information they acquired.

Part II: Interviews

Between June and July 2019, 22 men who completed a survey in the first phase of the study participated in semistructured interviews using Zoom videoconference software (Zoom software citation [53]). Only survey respondents who consented to being contacted to participate in an interview were considered for selection. Purposive sampling was used to select interviewees based on the following criteria: HIV-positive diagnosis (n=9), HIV-negative and using PrEP (n=7), and HIV-negative and not using PrEP (n=6). Given that most previous studies of HIV-related information behaviors have been conducted among PLWH, the purposive sampling criteria were developed to include the perspectives and experiences of YBMSM at different points of engagement along the HIV prevention and care continuum.

Interviews, which lasted on an average for 1 hour and 16 minutes, were guided by participant responses to survey items and included questions on topics including experiences of using mobile devices and the internet to locate HIV information, incidental acquisition of HIV information while using social networking sites and geospatial dating apps, experiences acquiring HIV information offline, motivators and deterrents of HIV information acquisition, the application of HIV information, use of technologies to overcome barriers to HIV information acquisition, and their criteria for evaluating and selecting HIV information sources. All participants who completed the interview were given a US $35 Amazon e–gift card. Interviews were audio recorded, transcribed verbatim, and member checked. Each participant was assigned a pseudonym.

The interview transcripts were analyzed thematically using Dedoose software (version 8.0.35) using grounded theory methodologies [54,55]. Our thematic analysis was guided by the Wilson model of information behavior, Erdelez model of information encountering, and Kari conceptualization of information use [51,52,56]. The Wilson model of information behavior explores the following:

> *The totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes face-to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given [51].*

The main constructs of the Wilson model of information behavior used for this thematic analysis were *activating mechanisms*, which are the factors motivating a decision to take action to satisfy a need for information (eg, information seeking), and the *sources and channels of information*. The Erdelez model of information encountering focuses on information acquired incidentally rather than through intentional information seeking [52]. We applied Kari conceptualization of information use as the application of a resource in some specific action [56].

Each interview transcript was read, and all texts related to HIV information acquisition and use were highlighted. The selected text was re-read and coded line-by-line to identify the initial emergent themes. Next, the most frequently reappearing initial codes were selected to begin explaining larger sections of data, and codes were condensed based on their thematic similarity. Codes were compared within and across interview text data to yield the most significant themes and provide more detailed information about HIV information acquisition and use among YBMSM.

XSL•FO
RenderX

Results

Internet Survey Sample

The average age of the internet sample was 29.2 (range 19-34) years. Most of the sample (70/83, 84%) consisted of men aged between 26 and 34 years. Most identified as gay (67/83, 80%), with 15% (13/83) identifying as bisexual. More than half of the sample (46/83, 55%) had less than a bachelor's degree and were employed (62/83, 74%; Table 1). Most participants had health insurance (56/83, 67%), and 32% (27/83) reported being uninsured. The annual household income varied with one-third (34%) of the participants earning less than US $20,000 per year. A total of 43% (36/83) of participants self-reported receiving an HIV-negative test result, 41% reported living with HIV, and 15% (13/83) were unsure of their HIV status. Of the men living with HIV, 68% (23/34) were using antiretroviral therapy (ART) at the time of the study, and only 30% (11/36) of the HIV-negative men used PrEP.

XSL•FO
RenderX

Table 1. Internet sample demographics (N=83).

Sample demographics	Value, n (%)
Age (years)[a]	
18-25	13 (15)
26-34	70 (84)
Education	
Less than high school	3 (3)
High school or GED[b]	35 (42)
Associate degree	8 (9)
Bachelor's degree	24 (29)
Graduate degree	13 (15)
Health insurance status	
Insured	56 (67)
Uninsured	27 (32)
Employment	
Unemployed	12 (14)
Full-time student	9 (11)
Employed part time	12 (14)
Employed full time	50 (60)
Annual household income (US $)	
<20,000	28 (34)
20,000-39,000	30 (36)
40,000-69,000	15 (18)
70,000-149,000	10 (12)
Residence	
Urban	53 (64)
Rural	18 (21)
Regional city and suburban	12 (14)
Sexual orientation	
Homosexual	67 (80)
Bisexual	13 (15)
Other	3 (3)
HIV testing, lifetime (self-report)	
Never tested	6 (7)
Previously tested	77 (93)
HIV status (self-report)	
Positive	34 (41)
Negative	36 (43)
Unknown	13 (15)
PrEP[c] use (self-report)	11 (13)
ART[d] use (self-report)	23 (28)

[a]Mean 29.2 (range 19-34) years.

[b]GED: general education development.

XSL•FO
RenderX

[c]PrEP: pre-exposure prophylaxis.

[d]ART: antiretroviral therapy.

Technology Ownership and Use

Participants reported significant access to multiple technological devices and frequent internet use (Table 2). Most participants (83/83, 100%) reported connecting to the internet through a cellular network on a mobile phone, with 94% (78/83) using a wireless connection. Approximately 100% (83/83) of the survey respondents who owned a mobile phone reported using the internet daily. In the web-based survey, the past use of eHealth and mHealth for HIV information was found to be high. Most men had used mobile phones or computers to seek HIV information.

Table 2. Technology ownership and use—internet sample (N=83).

Internet sample	Value, n (%)
Technology ownership	
Smartphone	79 (95)
Mobile phone	4 (5)
Laptop computer	63 (76)
Desktop computer	46 (43)
Tablet	47 (57)
Gaming console	58 (70)
Internet connectivity	
Cellular network (3G/4G)	83 (100)
Wi-Fi	78 (94)
Broadband via DSL[a], cable, or fiber optics	19 (23)
Frequency of internet use	
Use internet daily on smartphone or mobile phone	83 (100)
Past eHealth use and HIV information seeking	
Used mobile phone to look for HIV testing location	50 (60)
Used mobile phone to find other HIV information	79 (95)
Used computer to find HIV testing location	19 (23)
Used computer to find other HIV information	60 (72)

[a]DSL: digital subscriber line.

HIV Information Seeking and Source Selection

Most men (28/83, 34%) in the internet sample decided to seek HIV information as they believed it would be need-to-know information for the future (Table 3). Other participants (25/83, 30%) sought HIV information to satisfy their existing information needs and some (20/83, 24%) sought HIV information following a conversation with a member of their social network (eg, a friend, family member, or romantic partner). Approximately half of the participants (42/83, 50%) preferred to acquire HIV/AIDS information from a doctor or health care provider. The internet was the second most preferred source of HIV information at 36% (30/83). The 3 most trusted sources of HIV/AIDS information were a doctor or health care provider, the internet, and libraries, with 96% (80/83), 87% (72/83), and 72% (60/83), respectively, reporting that they trust these sources “a lot” or “some.” When asked, “the most recent time you looked for information related to HIV where did you go first?”, the internet was the top choice reported by 84% (70/83) of the participants. Web-based HIV information seeking was primarily mediated by default mobile phone search engines (eg, Google, Bing, and Yahoo). Consumer health information websites (31/83, 37%), government health information websites (30%), and social media sites (17%) were the top sources of intentionally sought web-based HIV information.

XSL•FO
RenderX

Table 3. HIV information acquisition and use—internet sample (N=83).

Internet sample	Value, n (%)
Decision to take action and seek HIV information	
Need-to-know information for the future	28 (34)
Satisfy existing information need	25 (30)
Following conversation with member of social network	20 (24)
Experiencing HIV or STI[a]-like symptoms	10 (12)
Selection of information source	
Preferred sources of HIV information	
Doctor or health care provider	42 (50)
Internet	30 (36)
Community organization	7 (8)
Friend or coworker	3 (3)
Print media (eg, book or pamphlet)	1 (1.4)
Trust on sources of HIV information	
Doctor or health care provider	80 (96)
Internet	72 (87)
Libraries	60 (72)
Newspaper or magazine	20 (24)
HIV information seeking	
Most recent source of HIV information	
Internet	70 (84)
Doctor or health care provider	9 (11)
Community organization	3 ()
Book	1 (1)
Sources of HIV information in the past 2 years	
Internet	80 (96)
Doctor or health care provider	41 (49)
Print media (eg, book, pamphlet, magazine, or newspaper)	29 (5)
Community organization	16 (19)
Friend or coworker	11 (13)
Sources of web-based HIV information	
Consumer health information	31 (37)
Government health information websites	25 (30)
Social media site	14 (17)
Community organization website	13 (16)
Sources of incidentally acquired HIV information	
Advertisements on social media sites and geospatial dating apps	54 (65)
Postings on social media sites from web-based social ties	44 (53)
Advertisements while browsing the web (non–social media platforms)	40 (48)
While talking to other people	30 (36)
While watching television or reading the news paper	27 (32)
Social media sites and geospatial dating apps (advertisements and postings from web-based social ties)	
Facebook	44 (53)

XSL•FO
RenderX

Internet sample	Value, n (%)
Instagram	40 (48)
Jack'd	38 (46)
Tinder	21 (25)
Grindr	20 (24)
Twitter	11 (13.2)
Tumblr	6 (7)
Adam4Adam	4 (5)
YouTube	2 (2.4)
Application of HIV information	
Find HIV or STI testing location	56 (67)
Received an HIV test	53 (64)
Begin treatment for HIV within 3 months following diagnosis (participants living with HIV)	19 (53)
Discuss HIV status with sexual partners	46 (55)

[a]STI: sexually transmitted infection.

Incidental HIV Information Acquisition

Incidental HIV information acquisition was high among the internet sample (Table 3) [57,58]. Respondents (54/83, 65%) most often incidentally acquired HIV information on the web via advertisements on social media sites and geospatial dating apps, postings on social media sites from their web-based social ties (eg, people and organizations they follow; 44/83, 53%), and advertisements encountered while browsing the internet (40/83, 48%). Most participants who incidentally acquired HIV/AIDS information while using social media sites or dating apps did so via Facebook (44/83, 53%), Instagram (40/83, 48%), and the dating app Jack'd (38/83, 46%).

Information Use

Most men in the internet sample applied the HIV information they acquired by looking for HIV or STI testing locations (56/83, 67%) and utilizing HIV or STI testing services (53/83, 64%; Table 3). Among participants who among participants who were living with HIV (34/83), more than half (19/34, 53%) used the knowledge of their status to begin ART within 3 months of receiving a diagnosis. Approximately 55% (46/83) of the internet sample used the HIV information they acquired to discuss HIV status with their sexual partners.

Interview Sample

Interview Sample Characteristics

Semistructured interviews were conducted with 22 YBMSM, aged between 22 and 33 years, who completed the web-based survey during the first phase of the study (Table 4). The interview sample included individuals residing in 9 counties in the state of North Carolina. A total of 9 men were living with HIV, 7 men were HIV-negative and using PrEP, and 6 men were HIV-negative and not using PrEP. The sample criteria were selected to reflect the varying HIV information needs and information interactions of YBMSM at different points of engagement along the HIV prevention and care continuum.

XSL•FO
RenderX

Table 4. Interview sample demographic characteristics (n=22).

Interview sample characteristics	Value, n (%)
Age (years)[a]	
18-25	5 (23)
26-34	17(77)
Education	
Less than high school	1 (4)
High school or GED[b]	7 (32)
Associate degree	3 (13)
Bachelor's degree	6 (27)
Graduate degree	5 (23)
Health insurance status	
Private health insurance	10 (46)
Medicaid	3 (13)
Uninsured	9 (41)
Employment	
Employed part time	6 (27)
Employed full time	11 (50)
Annual household income (US $)	
<20,000	9 (41)
20,000-39,000	8 (36)
40,000-69,000	5 (23)
Residence	
Urban	16 (73)
Rural	4 (18)
Regional city and suburban	2 (9)
Sexual orientation	
Homosexual	17 (77)
Bisexual	5 (23)
HIV testing, lifetime (self-report)	
Never tested	0
Previously tested	22 (100)
HIV status (self-report)	
Positive	9 (41)
Negative	13 (59)
PrEP[c] use (self-report)	7 (54)
ART[d] use (self-report)	9 (100)

[a]Mean 28.8 (range 22-34) years.

[b]GED: general education development.

[c]PrEP: pre-exposure prophylaxis.

[d]ART: antiretroviral therapy.

XSL•FO
RenderX

The Internet as a Primary Source of HIV Information

The use of the internet and mobile technology for HIV information was high in both the web and survey samples. All men in the interview sample confirmed using a mobile phone to search for HIV information on the internet. The most common themes were convenience and the ease of accessibility that mobile phones provided for seeking HIV information from multiple sources. Most perceived the HIV information they acquired on the web to be accurate:

> *Mostly because they usually give me the correct answer. It's really easy to navigate and you can find reputable sources like dot orgs or dot gov that's really accurate, something you can believe in as a source. That's why I trust it and then you don't have to go anywhere. Like you can pull out your phone and with quick searching, know the answer in like 10 seconds. Okay. Yep.* [Preston, HIV-negative, using PrEP, 23 years]

Government health information websites and health organization websites were deemed to be the most trustworthy sources of web-based HIV information. Men also valued the ability to seek HIV information privately through the internet while using a mobile phone but expressed concerns about others discovering their browsing history:

> *I usually use my phone and I like go into private browsing. You know like you just start typing something or if someone is using my phone I don't want them to start typing something, then it auto fills and then I hear "hey, what's this?"* [Nathan, HIV-positive, 30 years]

In the interviews, men confirmed that doctors and health care providers were their preferred and most trusted source of HIV information. Despite this preference, mobile HIV information seeking was used by men to circumvent barriers preventing them from accessing it from doctors and health care providers. The most common subthemes included discomfort communicating with health care providers about same-sex sexual behaviors and the underutilization of health care providers as an information source because of the experienced and anticipated intersectional stigma rooted in racism, homophobia, and HIV stigma in health care settings:

> *I didn't tell my doctor that I was sleeping with men, and he didn't ask. I'm already Black, and when they see me, they already treat me like I'm a statistic. I'd rather not talk to them about this stuff, because I don't want them to treat me even worse.* [Rashod, HIV-negative, not using PrEP, 22 years]

HIV Information Use

The internet was used to seek HIV information by men in the interview sample to assess their risk of acquiring or transmitting HIV, support their skill building for HIV prevention, inform patient–doctor communication, and learn about HIV prevention and treatment options (Table 5). HIV-negative and participants living with HIV used resources to build skills to discuss their HIV or STI status and testing history with their sexual partners. One participant described his experience looking for HIV information about PrEP and eventually initiating PrEP to minimize his HIV risk once he started dating someone who was HIV-positive:

> *My expartner was HIV-positive. And I knew that going into that relationship, so I started to do research myself. Just looking online, browsing different online websites. And then, when I wanted to pursue Truvada, I actually spoke with a local PRIDE center.* [Artez, HIV-negative, taking PrEP, 32 years]

Several participants used HIV information to inform patient-doctor communication, including (1) participants *self-diagnosing* themselves before seeing a health care provider, (2) verifying HIV information they acquired on the web by speaking with a health care provider, and (3) seeking information to educate health care providers about PrEP. One participant described his experience of self-diagnosing himself before speaking with a provider because of the presence of flu-like symptoms (Table 5).

XSL•FO
RenderX

Table 5. HIV information use—interview sample (n=22).

Interview sample	Illustrative quotes from participants
Manage or assess risk for HIV/AIDS	"When I learned about medication resistance, it made me really on top of taking my medication. Because previously, I might go in for a sore throat and they'll give you some antibiotics and they'll tell you, 'hey, make sure you take two a day for 10 days, make sure you finish it,' and here we are two months later, and I still have like 4 pills there. So, I wasn't always on top of it. Once I started feeling better, I'd be like, ok whatever. But learning about medication resistance particularly with HIV has def made me very much on top of taking medication as I've been directed to by my doctors. I have to aggressively handle my sexual health myself." [Jerry, HIV-positive, 28 years]
Inform patient-doctor communication	
Self-diagnosis	"Well, a lot of times I do a lot of self-diagnosing first just to kind of give them an idea as to what's going on, give myself an idea of what's going on. And to be able to explain my signs and symptoms more in detail. So that gives them a better view of what potentially could be going on. So, I went ahead and got tested just to make sure it was ok...and then if there was treatment to go along with it, it would be syphilis, the series of three shots, over the course of like six weeks...and I never wanted to go through that again, so I ask my doctor what to look for in myself and in others." [Chris, HIV-negative, non-PrEP[a] user, 29 years]
Information verification	"Whenever there is something new or a new medication or something that I hear about I will research it. And look at it. Or if it's like has to do with different studies, I'll look at the things that's part of the study. And then I will ask that information of my doctor." [Thomas, HIV-positive, 24 years]
Educate health care providers	"I just know I have a few friends that had experiences going to doctors, asking for PrEP or anything like that and the doctor was very closed minded or not very up-to-date on medical history in our aspects. We have to look for stuff on our own to tell them about it. I don't like that. I still mostly trust doctors, but I'll always think I'll have a shred of not doubt, but like I'm gonna double check and make sure you're telling me the best stuff. I'm going to make sure I'm with a good doctor...who is knowledgeable on whatever I need them to be knowledgeable on." [John, HIV-negative, PrEP user, 26 years]
Skill building for HIV prevention	"I started looking for this information because I needed help figuring out how to talk about it with other people. I'm to the point now where in the past if I got burned by somebody, I wouldn't ask them, I would just go get treated. I was afraid of saying something to this person. But I do better now just telling people like, 'hey I just went and got treated you might want to get tested on what's going on.' I was afraid to do so because the first thing you think about is them passing judgment." [Will, HIV-negative, non-PrEP user, 27 years]
Understand HIV prevention and treatment options	"I would like look up, cause I found out there's different strands to HIV, like it's not just one person has HIV, like there's levels to it. So, I'd try to Google and find information about that. And I also heard about PrEP, it was like maybe a cure or whatever I guess to try to prevent you from getting it, or lessen your chance of getting it. To be honest, my doctor had told me about it before, but I wanted to do my own research on it, so I just Googled and read up on that to find out what it's all about." [Patrick, HIV-negative, non-PrEP user, 32 years]

[a]PrEP: pre-exposure prophylaxis.

Use of Social Media for HIV Information Support Exchange

Although participants expressed displeasure with the framing and tone of advertisements on social media sites and geospatial dating apps promoting HIV prevention, they still used the sites as a space for informational support and communication with other Black MSM. Many participants reported acquiring HIV information from Black and other MSM of color who posted about their experiences living with HIV, being in a mixed HIV status relationship (serodiscordant), and the adoption of HIV risk-reduction behavior (eg, PrEP use and HIV self-testing):

> *I guess you can say that with me being on this medication, I do have a best friend that was also pretty much in the same boat as I am dealing with someone that was HIV positive as well. He really didn't know much about PrEP as well, so I talked to him about it, I gave him the same information that my doctor pretty much you know gave to me. I actually pretty much introduced him to my doctor that I'm with right now, and so they've been on PrEP for a while now. I guess you can say that I've been on it, and I hadn't received this information prior, and I was able to share it with someone else that was pretty much unaware of it as well.* [Antonio, HIV-negative, using PrEP, 27 years]

Unfortunately, some participants noted their diminished ability to exchange HIV information with Black and other MSM of color because of censorship policies on social media sites. When probed about where he obtained HIV information on the web, one participant noted his loss of HIV information from Black MSM on Tumblr. Other participants noted the loss of online community due to the suspension of some accounts and content removed because of censorship policies.

Concerns About Incidentally Acquired Web-Based HIV Information

The widespread use of social media sites and geospatial dating apps was confirmed in the interview sample. Encountering HIV information on these platforms was common. All participants expressed concerns about the framing and tone of the HIV prevention information and messages they encountered on the web via mass media HIV-prevention advertisements. Men described the framing of HIV prevention information as "judgmental" and perceived the tone of the information to be "pushy" and "fearmongering":

XSL•FO
RenderX

> *The ads were overreacting, but just in a way that's like, do this or you'll die or do this and you definitely get, you know, you'll definitely get syphilis...do this or you'll end up here or there and it's kind of stigmatizing, because I know I noticed a lot of their targeted audience where it wasn't really straight male, and it isn't fair to the LGBT community cause it's almost like, okay, you're targeting them, like making them feel like they have to get tested because they have something because they're a part of this community versus this guy, he's on the football team or this girl who is a cheerleader. Right. And they have, they may have something, not know it and passing it along to everyone else.* [Rick, HIV-negative, non-PrEP user, 22 years]

Discussion

Principal Findings

The purpose of this mixed methods study among YBMSM who use the internet was to explore their HIV-related health information behavior and to conduct formative work to inform the development of future, culturally tailored eHealth HIV interventions. The study found high technology ownership and use of mobile and social computing technologies to facilitate the utilization of HIV prevention and care services. Web-based HIV information was highly accessible to men in the sample through mobile phones and computers. Men in the interview sample discussed using eHealth to circumvent barriers to obtaining HIV information offline from a health care provider and for the ability to seek this information using private browsers. These findings indicate the importance of considering privacy and confidentiality when developing culturally tailored eHealth HIV interventions for YBMSM.

Men in the internet and interview sample preferred to acquire HIV information from a health care provider. However, men primarily used the internet to acquire HIV information because of barriers, including discomfort disclosing same-sex sexual behaviors to a health care provider, and intersecting stigmas experienced in health care settings. Previous studies with Black MSM have found perceptions of racism and medical mistrust to affect provider communication, health care access, and uptake of prevention services [59,60]. Black MSM have been found to be less likely to report their same-sex sexual behaviors to primary care providers compared with White MSM [57]. Prior research with young men who have sex with men (YMSM) reported barriers to communicating same-sex sexual behaviors and sexual health concerns to providers because of the fear of heterosexist bias, concerns about sexual health information being disclosed to parents, and a general belief that sexual minority youth did not receive equitable treatment in health care settings [58]. As reported in this study, the intersecting stigmas YBMSM experience related to sexuality, HIV, and racism can hinder access to HIV information from reliable and trusted sources of information [61,62]. To support engagement in HIV prevention and care, structural barriers to accessing HIV information such as racism, HIV-stigma, and homophobia must be minimized [63-65]. Future culturally tailored eHealth HIV interventions should aim to reduce intersectional stigma at the point of care and improve communication between providers and YBMSM. The development of these interventions should be community-driven. YBMSM need to be engaged in each step of the research process to identify and prioritize their needs and experiences during health care encounters. Health care providers should also receive training to provide antiracist, culturally competent care to YBMSM [66,67].

Health care providers must be equipped to provide timely recommendations on HIV prevention methods such as PrEP and HIV or STI testing [68]. In the interview sample, YBMSM shared experiences speaking with health care providers who were uninformed about PrEP and other HIV-prevention methods such as undetectable equals untransmittable (U=U; ie, a person who is HIV-positive with a consistently undetectable viral load cannot transmit HIV to a sex partner). To assist health care providers in providing timely HIV-prevention recommendations, the application of machine learning methods could be used to create an HIV risk prediction tool that identifies potential candidates for PrEP and other HIV-prevention strategies [69,70].

Previous research has found that when communicating about sensitive subjects, people may be more comfortable interacting with humans through a computer than through face-to-face interactions [71,72]. Owing to the high use of eHealth for HIV information acquisition and use among YBMSM in this sample, eHealth modalities may be especially advantageous for facilitating shared decision making between men and providers. eHealth interventions for YBMSM should aim to enhance shared decision making by increasing awareness and practical expectations of the advantages and disadvantages of prevention strategies such as PrEP and provide an environment in which they are comfortable communicating with providers [73,74]. This may help mitigate obstacles at the social, structural, and individual levels before enrolling them in health care services [75,76].

Concerns about the framing and tone of web-based, mass media HIV-prevention advertisements targeting YBMSM were raised during this study. Participants perceived the advertisements to be stigmatizing and judgmental of YBMSM. This aligns with research that found public health messaging aimed at reducing HIV disparities to be alienating [77-79]. Pleasure-centric and sex-positive framing of HIV information and messages may be helpful for promoting HIV risk-reduction behaviors among YBMSM [80]. To avoid further stigmatization of YBMSM, future eHealth HIV interventions should include community-informed, culturally tailored HIV messaging [81].

The use of social media sites for sharing and exchanging HIV information with web-based communities of Black MSM was a common theme of this study. Participants valued receiving HIV information from Black MSM who shared their experiences living with HIV, navigating serodiscordant relationships, and using PrEP and HIV self-testing kits. This aligns with the literature that found that social media sites are often used by LGBTQ individuals who may use these platforms to find information that may not be available through their offline social networks, formal education spaces, or health care providers [82,83].

XSL•FO
RenderX

In this study, it was noted that participants experienced diminished access to HIV information from Black MSM on the web because of censorship policies on social media sites such as Tumblr. These censorship policies are a direct result of FOSTA/SESTA legislation, and point to its harmful impact on minoritized communities who relied on these spaces for online social support building [84]. Participants described Tumblr as a platform where they built community and received nonstigmatizing information related to HIV. This finding supports previous research that conceptualized Tumblr as a "queer technology" and explored how the policy change pushed away communities of users, especially communities that relied on the "adult content" banned in this space for medical education and knowledge [85]. Future eHealth interventions should provide online spaces, such as forums, that promote community building and information support exchange among YBMSM.

Limitations

This study has several limitations. First, the findings of this study are limited in their generalizability and thus do not encompass the experiences of all YBMSM, as local and individual differences are also present from this data source. Many participants living with HIV in the study were recruited from HIV clinics and community-based organizations and thus may not be reflective of the population of YBMSM in North Carolina who are out of care. The size of the quantitative sample and qualitative nature of the study also make it difficult to generalize the results to all YBMSM in the United States. Future research would benefit from a larger study sample and a more rigorously recruited study sample. However, this study does show strength over previous internet-based research with YMSM, which has traditionally been limited to mostly White participants in a single geographic region for the sample. Second, internet access familiarity with web-based technology may have influenced the decision to participate in the study. YBMSM who use a different type of electronic device to access the internet may be able to navigate websites easily and may encounter technical challenges.

Although web-based surveys and video chat platforms for qualitative data collection have been found to reduce response bias and minimize barriers for research participation (eg, financial barriers related to travel expenses or time off from work, stigma associated with research participation, and physical disabilities precluding mobility), they do not account for limited literacy. Notwithstanding these limitations, our findings represent a novel characterization of HIV information acquisition and use among YBMSM in the southeastern United States who use the internet—a topic and population that warrants further study given the high rates of HIV diagnoses among this population in this geographic region. The data from this study may provide invaluable information concerning limitations to HIV information access and the potential to use technologies to reduce social, structural, and individual-level barriers to HIV prevention and treatment among YBMSM.

Acknowledgments

MT and KB were supported by the National Institute of Mental Health of the National Institutes of Health under award number R25MH087217. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries WL, Wilson PA, et al. Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis. The Lancet 2012 Jul;380(9839):341-348. [doi: 10.1016/S0140-6736(12)60899-X]
2. Centers for Disease Control & Prevention. HIV among African American gay and bisexual men. URL: http://www.cdc.gov/hiv/risk/racialethnic/bmsm/facts/ [accessed 2020-12-04]
3. Hess KL, Hu X, Lansky A, Mermin J, Hall HI. Lifetime risk of a diagnosis of HIV infection in the United States. Annals of Epidemiology 2017 Apr;27(4):238-243. [doi: 10.1016/j.annepidem.2017.02.003]
4. Matthews DD, Herrick AL, Coulter RWS, Friedman MR, Mills TC, Eaton LA, et al. Running Backwards: Consequences of Current HIV Incidence Rates for the Next Generation of Black MSM in the United States. AIDS Behav 2016 Jan;20(1):7-16 [FREE Full text] [doi: 10.1007/s10461-015-1158-z] [Medline: 26267251]
5. Hightow-Weidman L, LeGrand S, Choi SK, Egger J, Hurt CB, Muessig KE. Exploring the HIV continuum of care among young black MSM. PLoS ONE 2017 Jun 29;12(6):e0179688. [doi: 10.1371/journal.pone.0179688]
6. Maulsby C, Millett G, Lindsey K, Kelley R, Johnson K, Montoya D, et al. HIV among Black men who have sex with men (MSM) in the United States: a review of the literature. AIDS Behav 2014 Jan;18(1):10-25. [doi: 10.1007/s10461-013-0476-2] [Medline: 23620241]
7. Raifman J, Dean LT, Montgomery MC, Almonte A, Arrington-Sanders R, Stein MD, et al. Racial and Ethnic Disparities in HIV Pre-exposure Prophylaxis Awareness Among Men Who have Sex with Men. AIDS Behav 2019 Mar 14;23(10):2706-2709. [doi: 10.1007/s10461-019-02462-3]

XSL•FO
RenderX

8. Rebeiro PF, Ivey KS, Craig KS, Hulgan T, Huaman MA, Nash R, et al. New Faces of HIV Infection: Age, Race, and Timing of Entry into HIV Care in the Southeastern United States. Journal of the International Association of Providers of AIDS Care (JIAPAC) 2017 May 31;16(4):347-352. [doi: 10.1177/2325957417710719]
9. Eaton LA, Driffin DD, Bauermeister J, Smith H, Conway-Washington C. Minimal Awareness and Stalled Uptake of Pre-Exposure Prophylaxis (PrEP) Among at Risk, HIV-Negative, Black Men Who Have Sex with Men. AIDS Patient Care STDS 2015 Aug;29(8):423-429 [FREE Full text] [doi: 10.1089/apc.2014.0303] [Medline: 26083143]
10. Rouffiac A, Whiteley L, Brown L, Mena L, Craker L, Healy M, et al. A Mobile Intervention to Improve Uptake of Pre-Exposure Prophylaxis for Southern Black Men Who Have Sex With Men: Protocol for Intervention Development and Pilot Randomized Controlled Trial. JMIR Res Protoc 2020 Feb 20;9(2):e15781 [FREE Full text] [doi: 10.2196/15781] [Medline: 32130196]
11. Brooks RA, Landovitz RJ, Regan R, Lee S, Allen VC. Perceptions of and intentions to adopt HIV pre-exposure prophylaxis among black men who have sex with men in Los Angeles. Int J STD AIDS 2015 Jan 30;26(14):1040-1048. [doi: 10.1177/0956462415570159]
12. Bauermeister J, Meanley S, Pingel E, Soler J, Harper G. PrEP Awareness and Perceived Barriers Among Single Young Men who have Sex with Men. CHR 2014 Mar 31;11(7):520-527. [doi: 10.2174/1570162x12666140129100411]
13. Holloway IW, Tan D, Gildner JL, Beougher SC, Pulsipher C, Montoya JA, et al. Facilitators and Barriers to Pre-Exposure Prophylaxis Willingness Among Young Men Who Have Sex with Men Who Use Geosocial Networking Applications in California. AIDS Patient Care STDS 2017 Dec;31(12):517-527. [doi: 10.1089/apc.2017.0082] [Medline: 29211513]
14. Nelson LE, Ajiboye W, Djiadeu P, Odhiambo AJ, Pedersen C, Ramos SR, et al. A Web-Based Intervention to Reduce Decision Conflict Regarding HIV Pre-Exposure Prophylaxis: Protocol for a Clinical Trial. JMIR Res Protoc 2020 Jun 15;9(6):e15080. [doi: 10.2196/15080]
15. Kalichman SC, Picciano JF, Roffman RA. Motivation to Reduce HIV Risk Behaviors in the Context of the Information, Motivation and Behavioral Skills (IMB) Model of HIV Prevention. J Health Psychol 2008 Jul;13(5):680-689. [doi: 10.1177/1359105307082456]
16. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychology 2006;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462]
17. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Feb 12;37(3):173-191. [doi: 10.1111/hir.12287]
18. Fidel R. Human information interaction: an ecological approach to information behavior. In: MIT Press. Cambridge, MA: MIT Press; 2012.
19. Julien H, Fourie I. Reflections of affect in studies of information behavior in HIV/AIDS contexts: an exploratory quantitative content analysis. Libr Inf Sci Res 2015 Jan;37(1):3-9. [doi: 10.1016/j.lisr.2014.09.001]
20. Greyson D, Johnson J. The role of information in health behavior: a scoping study and discussion of major public health models. J Assn Inf Sci Tec 2015 Jun 24;67(12):2831-2841. [doi: 10.1002/asi.23392]
21. Hogan TP, Palmer CL. Information preferences and practices among people living with HIV/AIDS: results from a nationwide survey. J Med Libr Assoc 2005 Oct;93(4):431-439 [FREE Full text] [Medline: 16239938]
22. Huber JT, Cruz JM. Information Needs and Information-Seeking Behaviors of HIV Positive Men and Women. Medical Reference Services Quarterly 2000 Sep;19(3):39-48. [doi: 10.1300/j115v19n03_03]
23. Veinot TC, Meadowbrooke CC, Loveluck J, Hickok A, Bauermeister JA. How "community" matters for how people interact with information: mixed methods study of young men who have sex with other men. J Med Internet Res 2013 Feb 21;15(2):e33 [FREE Full text] [doi: 10.2196/jmir.2370] [Medline: 23428825]
24. 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]
25. Siegel K, Raveis V. Perceptions of Access to HIV-Related Information, Care, and Services among Infected Minority Men. Qual Health Res 2016 Jul;7(1):9-31. [doi: 10.1177/104973239700700102]
26. Samal L, Saha S, Chander G, Korthuis PT, Sharma RK, Sharp V, et al. Internet Health Information Seeking Behavior and Antiretroviral Adherence in Persons Living with HIV/AIDS. AIDS Patient Care and STDs 2011 Jul;25(7):445-449. [doi: 10.1089/apc.2011.0027]
27. Stonbraker S, Befus M, Lerebours Nadal L, Halpern M, Larson E. Factors Associated with Health Information Seeking, Processing, and Use Among HIV Positive Adults in the Dominican Republic. AIDS Behav 2016 Oct 6;21(6):1588-1600. [doi: 10.1007/s10461-016-1569-5]
28. Stevens R, Gilliard-Matthews S, Dunaev J, Todhunter-Reid A, Brawner B, Stewart J. Social Media Use and Sexual Risk Reduction Behavior Among Minority Youth. Nursing Research 2017;66(5):368-377. [doi: 10.1097/nnr.0000000000000237]
29. Dolcini MM, Warren J, Towner SL, Catania JA, Harper GW. Information Age: Do Urban African American Youth Find Sexual Health Information Online? Sex Res Social Policy 2015 Jun 01;12(2):110-114 [FREE Full text] [doi: 10.1007/s13178-014-0174-5] [Medline: 26052264]
30. Jones RK, Biddlecom AE. Is the Internet Filling the Sexual Health Information Gap for Teens? An Exploratory Study. Journal of Health Communication 2011 Jan 31;16(2):112-123. [doi: 10.1080/10810730.2010.535112]

XSL•FO
RenderX

31. Namuleme RK. HIV/AIDS-related stigma and information behaviour: an ethnographic study in the UK. Health Info Libr J 2015 Feb 13;32(1):61-66. [doi: 10.1111/hir.12090]
32. 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 Dec 13;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]
33. Fields EL, Long A, Dangerfield DT, Morgan A, Uzzi M, Arrington-Sanders R, et al. There's an App for That: Using Geosocial Networking Apps to Access Young Black Gay, Bisexual, and other MSM at Risk for HIV. Am J Health Promot 2019 Jul 30;34(1):42-51. [doi: 10.1177/0890117119865112]
34. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Carey MP. Mobile technology use and desired technology-based intervention characteristics among HIV+ Black men who have sex with men. AIDS Care 2016 Aug 18;29(4):423-427. [doi: 10.1080/09540121.2016.1220479]
35. Tan JYR, Nguyen TT, Tabrisky A, Siedle-Khan R, Napoles AM. Mobile Technology for Healthy Aging Among Older HIV-Positive Black Men Who Have Sex with Men: Qualitative Study. JMIR Aging 2018 Dec 10;1(2):e11723. [doi: 10.2196/11723]
36. van den Berg JJ, Silverman T, Fernandez MI, Henny KD, Gaul ZJ, Sutton MY, et al. Using eHealth to Reach Black and Hispanic Men Who Have Sex With Men Regarding Treatment as Prevention and Preexposure Prophylaxis: Protocol for a Small Randomized Controlled Trial. JMIR Res Protoc 2018 Jul 16;7(7):e11047. [doi: 10.2196/11047]
37. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. mHealth 2020 Jan;6:7-7. [doi: 10.21037/mhealth.2019.10.01]
38. Hightow-Weidman L, Muessig K, Claude K, Roberts J, Zlotorzynska M, Sanchez T. Maximizing Digital Interventions for Youth in the Midst of Covid-19: Lessons from the Adolescent Trials Network for HIV Interventions. AIDS Behav 2020 Apr 18;24(8):2239-2243. [doi: 10.1007/s10461-020-02870-w]
39. Turner D, Lockhart E, Marhefka SL. Willingness of MSM Living with HIV to Take Part in Video-Groups: Application of the Technology Readiness and Acceptance Model. AIDS Behav 2019 May 23;23(11):3165-3174. [doi: 10.1007/s10461-019-02534-4]
40. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441. [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
41. Khanna AS, Schumm P, Schneider JA. Facebook network structure and awareness of preexposure prophylaxis among young men who have sex with men. Ann Epidemiol 2017 Mar;27(3):176-180 [FREE Full text] [doi: 10.1016/j.annepidem.2016.11.006] [Medline: 28003117]
42. Muessig KE, Baltierra NB, Pike EC, LeGrand S, Hightow-Weidman LB. Achieving HIV risk reduction through HealthMpowerment.org, a user-driven eHealth intervention for young Black men who have sex with men and transgender women who have sex with men. Digit Cult Educ 2014;6(3):164-182 [FREE Full text] [Medline: 25593616]
43. Barry MC, Threats M, Blackburn NA, LeGrand S, Dong W, Pulley DV, et al. "Stay strong! keep ya head up! move on! it gets better!!!!": resilience processes in the healthMpowerment online intervention of young black gay, bisexual and other men who have sex with men. AIDS Care 2019 Jan 11;30(sup5):S27-S38. [doi: 10.1080/09540121.2018.1510106]
44. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A Multilevel Tailored Web App–Based Intervention for Linking Young Men Who Have Sex With Men to Quality Care (Get Connected): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 Aug 02;7(8):e10444. [doi: 10.2196/10444]
45. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A Randomized Trial of an Online Risk Reduction Intervention for Young Black MSM. AIDS Behav 2019 May;23(5):1166-1177. [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
46. Blackburn NA, Dong W, Threats M, Barry M, LeGrand S, Hightow-Weidman LB, et al. Building community in the HIV online intervention space: lessons from the HealthMPowerment Intervention. Health Educ Behav 2021 Apr 09:- [FREE Full text] [doi: 10.1177/10901981211003859] [Medline: 33834892]
47. Threats M, Boyd DT, Diaz JE, Adebayo OW. Deterrents and motivators of HIV testing among young Black men who have sex with men in North Carolina. AIDS Care 2020 Nov 27:1-9. [doi: 10.1080/09540121.2020.1852161] [Medline: 33244990]
48. Ivankova NV. Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. Field Methods 2006 Feb 01;18(1):3-20. [doi: 10.1177/1525822X05282260]
49. Stonbraker S, Smaldone A, Luft H, Cushman LF, Lerebours Nadal L, Halpern M, et al. Associations between health literacy, HIV-related knowledge, and information behavior among persons living with HIV in the Dominican Republic. Public Health Nurs 2017 Dec 29;35(3):166-175. [doi: 10.1111/phn.12382]
50. Meadowbrooke CC, Veinot TC, Loveluck J, Hickok A, Bauermeister JA. Information behavior and HIV testing intentions among young men at risk for HIV/AIDS. J Assn Inf Sci Tec 2014 Jan 15;65(3):609-620. [doi: 10.1002/asi.23001]
51. Wilson T. Information behaviour: An interdisciplinary perspective. Information Processing & Management 1997 Jul;33(4):551-572. [doi: 10.1016/S0306-4573(97)00028-9]
52. Erdelez S. Investigation of information encountering in the controlled research environment. Information Processing & Management 2004 Nov;40(6):1013-1025. [doi: 10.1016/j.ipm.2004.02.002]

XSL•FO
RenderX

53. Zoom download center. URL: https://zoom.us/download [accessed 2021-04-29]
54. Charmaz K. Constructing grounded theory. London: SAGE; 2014.
55. Rubin H, Rubin I. Qualitative interviewing: the art of hearing data. Los Angeles: SAGE; 2012.
56. Kari J. Diversity in the conceptions of information use. Information research 2010;15(3):--.
57. Petroll AE, Mosack KE. Physician awareness of sexual orientation and preventive health recommendations to men who have sex with men. Sex Transm Dis 2011 Jan;38(1):63-67 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181ebd50f] [Medline: 20706178]
58. Fisher CB, Fried AL, Macapagal K, Mustanski B. Patient–Provider Communication Barriers and Facilitators to HIV and STI Preventive Services for Adolescent MSM. AIDS Behav 2018 Mar 15;22(10):3417-3428. [doi: 10.1007/s10461-018-2081-x]
59. Hausmann LRM, Hannon MJ, Kresevic DM, Hanusa BH, Kwoh CK, Ibrahim SA. Impact of Perceived Discrimination in Healthcare on Patient-Provider Communication. Medical Care 2011;49(7):626-633. [doi: 10.1097/mlr.0b013e318215d93c]
60. Eaton LA, Driffin DD, Smith H, Conway-Washington C, White D, Cherry C. Psychosocial factors related to willingness to use pre-exposure prophylaxis for HIV prevention among Black men who have sex with men attending a community event. Sex. Health 2014;11(3):244. [doi: 10.1071/sh14022]
61. Eaton LA, Driffin DD, Kegler C, Smith H, Conway-Washington C, White D, et al. The Role of Stigma and Medical Mistrust in the Routine Health Care Engagement of Black Men Who Have Sex With Men. Am J Public Health 2015 Feb;105(2):e75-e82. [doi: 10.2105/ajph.2014.302322]
62. Threats M, Boyd D, Diaz J, Adebayo O. Deterrents and motivators of HIV testing among young Black men who have sex with men in North Carolina. AIDS Care 2020:1-9 [FREE Full text] [doi: 10.1080/09540121.2020.1852161] [Medline: 33244990]
63. Pohl A, Trill R. Digital Health Literacy as Precondition for Sustainable and Equal Health Care - A Study Focusing the Users' Prescriptive. In: Li H, Nykänen P, Suomi R, Wickramasinghe N, Widén G, Zhan M, editors. Building Sustainable Health Ecosystems. WIS 2016. Communications in Computer and Information Science. New York: Springer, Cham; 2016:37-46.
64. Gibson AN, Martin JD. Re‐situating information poverty: Information marginalization and parents of individuals with disabilities. Journal of the Association for Information Science and Technology 2019 Feb 11;70(5):476-487. [doi: 10.1002/asi.24128]
65. Threats M, Brawner BM, Montgomery TM, Abrams J, Jemmott LS, Crouch PC, et al. A review of recent HIV prevention interventions and future considerations for nursing science. J Assoc Nurses AIDS Care 2021;32(3):373-391. [doi: 10.1097/JNC.0000000000000246] [Medline: 33929980]
66. Quinn K, Dickson-Gomez J, Zarwell M, Pearson B, Lewis M. “A Gay Man and a Doctor are Just like, a Recipe for Destruction”: How Racism and Homonegativity in Healthcare Settings Influence PrEP Uptake Among Young Black MSM. AIDS Behav 2018 Dec 18;23(7):1951-1963. [doi: 10.1007/s10461-018-2375-z]
67. Arrington-Sanders R, Hailey-Fair K, Wirtz AL, Morgan A, Brooks D, Castillo M, et al. Role of Structural Marginalization, HIV Stigma, and Mistrust on HIV Prevention and Treatment Among Young Black Latinx Men Who Have Sex with Men and Transgender Women: Perspectives from Youth Service Providers. AIDS Patient Care and STDs 2020 Jan 01;34(1):7-15. [doi: 10.1089/apc.2019.0165]
68. Mutchler MG, McDavitt B, Ghani MA, Nogg K, Winder TJ, Soto JK. Getting PrEPared for HIV Prevention Navigation: Young Black Gay Men Talk About HIV Prevention in the Biomedical Era. AIDS Patient Care and STDs 2015 Sep;29(9):490-502. [doi: 10.1089/apc.2015.0002]
69. Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. The Lancet HIV 2019 Oct;6(10):e688-e695. [doi: 10.1016/s2352-3018(19)30137-7]
70. Gilkey MB, Marcus JL, Garrell JM, Powell VE, Maloney KM, Krakower DS. Using HIV Risk Prediction Tools to Identify Candidates for Pre-Exposure Prophylaxis: Perspectives from Patients and Primary Care Providers. AIDS Patient Care and STDs 2019 Aug;33(8):372-378. [doi: 10.1089/apc.2019.0056]
71. Dinoia J, Schinke S, Pena J, Schwinn T. Evaluation of a brief computer-mediated intervention to reduce HIV risk among early adolescent females. Journal of Adolescent Health 2004 Jul;35(1):62-64. [doi: 10.1016/s1054-139x(03)00371-9]
72. Evans A, Edmundson-Drane E, Harris K. Computer-assisted instruction: an effective instructional method for HIV prevention education? Journal of Adolescent Health 2000 Apr;26(4):244-251 [FREE Full text] [doi: 10.1016/s1054-139x(99)00093-2]
73. Stacey D, Hawker G, Dervin G, Tugwell P, Boland L, Pomey M, et al. Decision aid for patients considering total knee arthroplasty with preference report for surgeons: a pilot randomized controlled trial. BMC Musculoskelet Disord 2014 Feb 24;15(1):1-10 [FREE Full text] [doi: 10.1186/1471-2474-15-54] [Medline: 24564877]
74. Jacklin S, Maskrey N, Chapman S. Improving Shared Decision Making Between Patients and Clinicians: Design and Development of a Virtual Patient Simulation Tool. JMIR Med Educ 2018 Nov 06;4(2):e10088 [FREE Full text] [doi: 10.2196/10088] [Medline: 30401667]

XSL•FO
RenderX

75. Anand T, Nitpolprasert C, Ananworanich J, Pakam C, Nonenoy S, Jantarapakde J, et al. Innovative strategies using communications technologies to engage gay men and other men who have sex with men into early HIV testing and treatment in Thailand. Journal of Virus Eradication 2015 Apr;1(2):111-115. [doi: 10.1016/s2055-6640(20)30483-0]
76. Anand T, Nitpolprasert C, Trachunthong D, Kerr SJ, Janyam S, Linjongrat D, et al. A novel Online-to-Offline (O2O) model for pre-exposure prophylaxis and HIV testing scale up. Journal of the International AIDS Society 2017 Jan 01;20(1):21326. [doi: 10.7448/ias.20.1.21326]
77. Drumhiller K, Murray A, Gaul Z, Aholou TM, Sutton MY, Nanin J. "We Deserve Better!": Perceptions of HIV Testing Campaigns Among Black and Latino MSM in New York City. Arch Sex Behav 2017 Apr 25;47(1):289-297. [doi: 10.1007/s10508-017-0950-4]
78. Johnny L, Mitchell C. "Live and Let Live": An Analysis of HIV/AIDS-Related Stigma and Discrimination in International Campaign Posters. Journal of Health Communication 2006 Sep;11(8):755-767. [doi: 10.1080/10810730600934708]
79. Spieldenner AR, Castro CF. Education and Fear: Black and Gay in the Public Sphere of HIV Prevention. Communication Education 2010 Jul;59(3):274-281. [doi: 10.1080/03634521003606202]
80. Dehlin JM, Stillwagon R, Pickett J, Keene L, Schneider JA. #PrEP4Love: An Evaluation of a Sex-Positive HIV Prevention Campaign. JMIR Public Health Surveill 2019 Jun 17;5(2):e12822. [doi: 10.2196/12822]
81. Kerr J, Ayangeakaa S, Combs R, Harris L, Sears J, Northington T, et al. Community-Informed Development of a Campaign to Increase HIV Pre-exposure Prophylaxis (PrEP) Awareness Among African-American Young Adults. J Racial Ethn Health Disparities 2020 Aug 31:1-11. [doi: 10.1007/s40615-020-00848-x] [Medline: 32869211]
82. Cho A. Default publicness: Queer youth of color, social media, and being outed by the machine. New Media & Society 2017 Dec 12;20(9):3183-3200. [doi: 10.1177/1461444817744784]
83. Duguay S. Why tumblr's ban on adult content is bad for LGBTQ youth.: Phys.org URL: https://phys.org/news/2018-12-tumblr-adult-content-bad-lgbtq.html [accessed 2020-07-29]
84. Sawicki DA, Meffert BN, Read K, Heinz AJ. Culturally competent health care for sex workers: an examination of myths that stigmatize sex-work and hinder access to care. Sex Relation Ther 2019;34(3):355-371 [FREE Full text] [doi: 10.1080/14681994.2019.1574970] [Medline: 30899197]
85. Haimson OL, Dame-Griff A, Capello E, Richter Z. Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies. Feminist Media Studies 2019 Oct 18:1-17. [doi: 10.1080/14680777.2019.1678505]

Abbreviations

ART: antiretroviral therapy
LGBTQ: lesbian, gay, bisexual, trans, and queer
mHealth: mobile health
MSM: men who have sex with men
PLWH: people living with HIV
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection
YBMSM: young Black men who have sex with men
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 29.07.20; peer-reviewed by D Boyd, H Cho, A Khanna; comments to author 02.10.20; revised version received 31.10.20; accepted 16.01.21; published 07.05.21.

Please cite as:
Threats M, Bond K
HIV Information Acquisition and Use Among Young Black Men Who Have Sex With Men Who Use the Internet: Mixed Methods Study
J Med Internet Res 2021;23(5):e22986
URL: https://www.jmir.org/2021/5/e22986
doi:10.2196/22986
PMID:33960953

©Megan Threats, Keosha Bond. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.05.2021. 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

Validation of 4D Components for Measuring Quality of the Public Health Data Collection Process: Elicitation Study

Hong Chen[1,2], MPH; Ping Yu[1,3], PhD; David Hailey[1], PhD; Tingru Cui[4], PhD

[1]School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia
[2]Jiangxi Provincial Centre for Disease Control and Prevention, Nanchang, China
[3]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia
[4]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

Corresponding Author:
Ping Yu, PhD
School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 2 42215412
Fax: 61 2 4221 4170
Email: ping@uow.edu.au

Abstract

Background: Identification of the essential components of the quality of the data collection process is the starting point for designing effective data quality management strategies for public health information systems. An inductive analysis of the global literature on the quality of the public health data collection process has led to the formation of a preliminary 4D component framework, that is, data collection management, data collection personnel, data collection system, and data collection environment. It is necessary to empirically validate the framework for its use in future research and practice.

Objective: This study aims to obtain empirical evidence to confirm the components of the framework and, if needed, to further develop this framework.

Methods: Expert elicitation was used to evaluate the preliminary framework in the context of the Chinese National HIV/AIDS Comprehensive Response Information Management System. The research processes included the development of an interview guide and data collection form, data collection, and analysis. A total of 3 public health administrators, 15 public health workers, and 10 health care practitioners participated in the elicitation session. A framework qualitative data analysis approach and a quantitative comparative analysis were followed to elicit themes from the interview transcripts and to map them to the elements of the preliminary 4D framework.

Results: A total of 302 codes were extracted from interview transcripts. After iterative and recursive comparison, classification, and mapping, 46 new indicators emerged; 24.8% (37/149) of the original indicators were deleted because of a lack of evidence support and another 28.2% (42/149) were merged. The validated 4D component framework consists of 116 indicators (82 facilitators and 34 barriers). The first component, data collection management, includes data collection protocols and quality assurance. It was measured by 41 indicators, decreased from the original 49% (73/149) to 35.3% (41/116). The second component, data collection environment, was measured by 37 indicators, increased from the original 13.4% (20/149) to 31.9% (37/116). It comprised leadership, training, funding, organizational policy, high-level management support, and collaboration among parallel organizations. The third component, data collection personnel, includes the perception of data collection, skills and competence, communication, and staffing patterns. There was no change in the proportion for data collection personnel (19.5% vs 19.0%), although the number of its indicators was reduced from 29 to 22. The fourth component, the data collection system, was measured using 16 indicators, with a slight decrease in percentage points from 18.1% (27/149) to 13.8% (16/116). It comprised functions, system integration, technical support, and data collection devices.

Conclusions: This expert elicitation study validated and improved the 4D framework. The framework can be useful in developing a questionnaire survey instrument for measuring the quality of the public health data collection process after validation of psychometric properties and item reduction.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e17240) doi:10.2196/17240

KEYWORDS

data quality; data collection; HIV/AIDS; public health informatics; health information systems; component validation; expert elicitation; public health; health informatics

Introduction

Background

Public health, a data-intensive discipline, relies on high-quality data to monitor the health and well-being of the population, make appropriate policy decisions for intervention, and evaluate intervention outcomes [1-3]. After two decades of development in the design and implementation of information and communication technologies (ICTs) for national public health data management, public health information systems (PHISs) have developed into essential data repositories [1,4,5]. PHISs have been well integrated into many nations' health information management systems, such as those of the United States, Australia, and China [6-9]. The data stored in PHISs, for example, on women's and children's health, aging population, and people living with HIV/AIDS, have enabled public health agencies to formulate evidence-based policies and plan and evaluate program performance to ensure accountability for citizens and countries [1,6,7,10].

As data-driven public health management assumes data are accurate, timely, and reliable, data quality assessment needs to be continuously and rigorously conducted to ensure high-quality data in PHISs [4]. Data quality is a 3D concept that includes the quality of data, data collection process, and data use. Improving the quality of the data collection process is critical for PHIS data quality management [11]. Identification of the essential components of the quality of the PHIS data collection process is the starting point for the design of effective public health data quality management strategies [4,7].

Through appraisal and synthesis of literature that reports the factors affecting the rigor of the PHIS data collection process, we have proposed a preliminary conceptual framework that focuses on four dimensions of the quality of the process [12]. These are data collection management, data collector, information system, and the data collection environment. We name them 4D components, which consist of 12 subcomponents and 149 indicators (Multimedia Appendix 1 [7,12-15]). Data collection management is an administrative process by which data are acquired, validated, stored, protected, and processed [7,13]. Its indicators include appropriate data collection methods, data entry forms, and ongoing quality assurance. At the individual level, data collection personnel (replacing *data collector*) need to have a right attitude, adequate skills, and competence for the job. They must maintain adequate communication with each other. For them to execute their tasks adequately, their organization needs to provide adequate staffing with the right skill mix [12]. A data collection system (replacing *information system*) requires different systems and elements to integrate and assist data capture, data entry, and data logging. Thus, continuous and systemic functionality and technical support are required [14]. A good data collection environment includes training, strong leadership, and funding support for data collection [15]. Given that this preliminary 4D component framework was derived from an inductive analysis of the literature, validation of the framework within a certain PHIS was needed.

Expert elicitation is a research method used to identify and address uncertain subjects, especially when relevant local evidence or information is incomplete [16]. This method has been widely used in public health for policy decisions to generate evidence [17,18] to achieve various research goals, such as environmental health impact assessment [16], health technology assessment [19], and economic evaluation of health gains of antenatal care [20]. Knowledge synthesized from expert opinions can form the foundation for further research.

The main procedures for a formal expert elicitation include characterization of uncertainties, selection of experts, elicitation of expert judgments, and possible aggregation and reporting in a temporary summary [16]. The criteria for expert selection include the following: the person should be representative of the main population of interest and he or she should have sufficient intellectual ability to provide the theoretical definitions, rank the importance of the data items, and explain a potential causal relationship between them [16]. Expert judgments should adhere to the principles of the scientific process. These are accountability, neutrality, fairness, and the ability for empirical control [21]. A facilitator, often a trained interviewer, has *the enormous potentialto reduce bias in expert elicitation* by clarifying the questions to the expert [16,19]. A systematic elicitation session could increase the validity, transparency, and trustworthiness of research [16].

Objectives

Using an expert elicitation approach, this study aims to obtain empirical evidence to confirm the components of the 4D framework and, if needed, to further develop the framework.

Methods

Study Setting

The study was conducted within a country-level PHIS, the Chinese HIV/AIDS Comprehensive Response Information Management System (CRIMS). Acknowledged as one of the milestones for China's HIV/AIDS response programs over the past three decades [22], the CRIMS is a web-based national AIDS information management system that was established in 2008 [5]. The system has been used for routine HIV/AIDS prevention and control data collection from hospitals and all units of the Chinese Center for Disease Control and Prevention (China CDC) in 2893 counties. The data stored in the CRIMS include demographic information, case reporting, antiretroviral treatment, methadone maintenance therapy, behavioral interventions, laboratory testing, counseling, and surveillance. These real-time data can be used for decision making,

XSL•FO
RenderX

monitoring, and evaluating HIV/AIDS prevention and control programs in health bureaus and CDCs at national, provincial, city, and county levels [10]. Therefore, high-quality data in the system are imperative for China's HIV/AIDS program monitoring and evaluation. However, prior studies found that public health professionals lacked trust in the quality of data in the CRIMS and expressed concerns over the quality of the data collection process [17,23,24]. This primary concern of public health professionals in China has also motivated this study.

Data management within the CRIMS includes data collection, data entry, data analysis, data assurance, and data use [25]. The personnel involved in the CRIMS data management include health administrators in health bureaus, epidemiologists and laboratory technicians in CDCs, and clinicians and data registrars in hospitals. They have accumulated rich experiences from long-term empirical work in HIV/AIDS data management and were thus appropriate experts who could provide inputs for this study.

Ethical Approval

This study was approved by the Human Research Ethics Committee at the University of Wollongong and the Institutional Review Board of the National Center for AIDS/STD Control and Prevention at the China CDC. All participants provided informed written consent to participate in the study and to publish individual data.

Design of Interview Guide and Data Collection Form

To ensure the validity of the study, we followed three broad categories of validity for qualitative research in information systems proposed by Venkatesh [26]. These are (1) design validity (eg, descriptive validity, credibility, and transferability), (2) analytical validity (eg, theoretical validity, dependability, consistency, and plausibility), and (3) inferential validity (eg, interpretive validity and confirmability).

During the design phase, an interview guide was developed in consultation with 7 information system researchers at the University of Wollongong: a professor, an associate professor, a lecturer, a research assistant, and 3 PhD candidates. Two open-ended questions were suggested: "What are the components of quality of the CRIMS data collection process?" and "What are the attributes of these components?"

An item represents a component or subcomponent of the 4D component framework in reference to the literature [12]. An item weight table was developed to elicit an expert's opinion about whether an item is a component or subcomponent of the quality of the CRIMS data collection process. To avoid bias in directing the expert to the preliminary 4D component framework, we reconstructed the testing items according to group discussions with consultant researchers. Four items that are not part of the framework but frequently identified in practice were added, including parallel organization, high-level management, social factors, and organizational policy. Four items that are elements of a certain original subcomponent or component were used to represent their parental components. These were data collection forms and data quality assessment strategies of the component data collection management, data collector's data quality audit skills, and demographics of the component data collection personnel. Four original items—funding, data collection personnel's communication, staffing pattern, and integration of different systems—were purposely excluded to test the completeness of the framework item spectrum. Each item was answered as *yes* or *no*. If the answer was yes, the expert was asked to rank the importance of the item for the quality of the CRIMS data collection process. The rankings ranged from 1 (the highest contribution) to 16 (the smallest contribution; Multimedia Appendix 2).

The interview guide and item weight table were translated into Chinese. Three bilingual authors validated the Chinese translation. The interview guide was pilot tested for content validity and face validity with 8 Chinese public health practitioners who worked within the CRIMS. All 8 practitioners agreed with the fit of the interview questions and the item weight table for the study.

Sampling and Recruitment of Study Participants

To ensure generalizability of the study, personnel from all administrative levels in all types of organizations with at least one data management role for the CRIMS were considered as potential experts. They were eligible for inclusion as experts to ensure a comprehensive capture of diverse expert opinions. Those who did not have a role in CRIMS data management were excluded.

Following the aforementioned selection criteria, we used a stratified sampling method to identify the participating organizations [27]. Representativeness was ensured by a lack of significant statistical differences in data quality among provinces [23,24]. A total of 19 organizations were selected, including 3 departments of health bureaus (1 each at the central, provincial, and county levels), 10 departments of the CDCs (1 each at the national, provincial, and city levels and 7 at the county level), and 6 hospitals (4 tertiary, 1 secondary, and 1 primary).

HC was an epidemiologist who specialized in HIV/AIDS prevention and control in a provincial CDC in China. She used a convenient sampling method to recruit participants working in health bureaus and CDCs. A personalized invitation message containing a cover letter and a consent form was sent through the Chinese social media QQ to 20 potential participants. All participants agreed to participate by returning a completed consent form. Web-based interviews were arranged with 18 of them through QQ media. The other 2 withdrew quoting time constraints. Of the 18 participants, 3 were from health bureaus at 3 different levels. The remaining 15 came from 4 tiers of the CDCs: 1 at the national level, 4 at the provincial level, 3 at the city level, and 7 at the county level.

HC recruited potential participants from 6 hospitals via direct contact with hospital management. She explained the project's purpose and research process to the relevant managers in the hospitals and sought their support in recommending eligible data management personnel to participate in the field study. Being introduced by the facility management, HC contacted the potential participant and organized an interview with the person at a designated venue and time. An interview would start only after providing written consent. Overall, 10 potential

XSL•FO
RenderX

participants were recommended and completed interviews. Of the 10 participants, 6 came from 4 tertiary hospitals, 3 from a secondary hospital, and the other from a primary health care center.

On average, the 28 participants had worked in public health or health services for 12 (SD 7) years and in the HIV/AIDS domain for 7 (SD 4) years. Of the 28 participants, 16 (57%) were female; 23 (82%) participants were aged between 30 and 50 years, and the other 5 (18%) were aged under 30 years. Most participants (25/28, 89%) had multiple job roles in HIV/AIDS data management.

Interview Procedure

Both telephone and face-to-face interviews were conducted. An internet voice call was made for telephone interviews with the practitioners during their work break or after hours. Face-to-face interviews were conducted at hospitals. The average duration of the interviews was 44 minutes (SD 23 min).

Each interview started with asking the practitioner to provide answers to the 2 open-ended questions. Answers from 3 of the first 5 practitioners were related to data quality instead of the focused topic of this study, the quality of the data collection process. To clarify the research topic, the researcher started subsequent interviews with the question, “What do you think the differences are between data quality and quality of the CRIMS data collection process?” A further probe clarified any emerging issues raised by the practitioners. Once information saturation was reached, that is, no further issues emerged, the interview was concluded.

After the practitioners answered all the open-ended questions, they were invited to assess the 16 items listed in the weight table. The other 7 items were raised by the practitioners. Their average rankings were not calculated because of the small sample size.

Data Processing and Analysis

All audio recordings were transcribed verbatim. The transcripts were sent to the interviewees for confirmation, translated into English, and back translated. Qualitative data analysis was conducted in accordance with the framework analysis approach suggested by Pope et al [28]. The theoretical (thematic) framework was the 4D component of the quality of the PHIS data collection process. The unit of analysis was each transcript. The data analysis was conducted in 3 steps.

Step 1: Generating the Initial Codes

Each transcript was thoroughly read to identify and understand the meaning of the relevant text. A concise phrase was created to summarize a sentence. For example, “Reward and punishment system, which is important for a working system...This should be in organizational management policy” (C102) was coded as “clear reward and punishment in organizational policy.” “If they (managers) understand the importance to the job (data collection), you will work easily; if they don’t, it is hard” (H306) were coded as “managers should understand the importance of data collection.”

After the first round of transcript encoding, 302 codes were extracted and stored in an Excel database.

Step 2: Mapping the Codes Using the Preliminary 4D Component Framework

All the 302 codes were compared and mapped with the original indicators and subcomponents in the preliminary framework. Three data processing strategies were used in 3 different scenarios.

Scenario 1

When a code had a similar meaning to an original subcomponent or indicator of the preliminary 4D framework, the original subcomponent or indicator remained or was further refined by merging, condensing, and grouping to represent the code.

Scenario 2

When the meaning of a code was not matched by any original subcomponent or indicator in the preliminary 4D framework, a judgment was made to add the code as a new subcomponent or indicator to the framework.

Scenario 3

When no empirical data could match the meaning of a certain subcomponent or indicator in the preliminary 4D framework, the subcomponent or indicator was deleted from the framework.

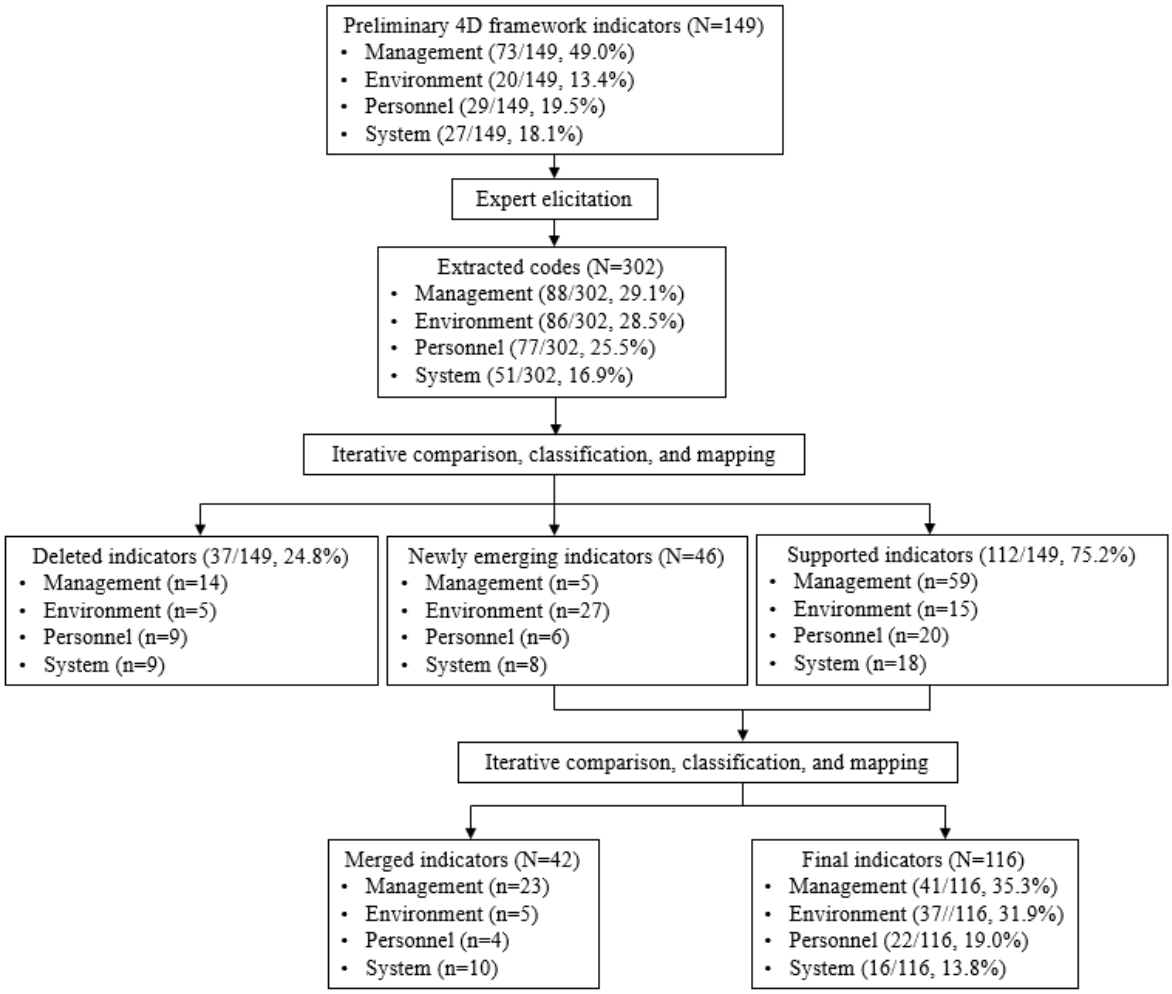
Iterative and recursive coding, mapping, and classification processes were applied continuously between steps 1 and 2. The 302 codes converged to the 4D component framework; 88 were grouped into the component data collection management, 86 into the data collection environment, 77 into the data collection personnel, and the remaining 51 into the data collection system. A total of 46 new indicators emerged from the extracted codes. Of the 149 original indicators, 37 (24.8%) were deleted because of a lack of evidence support and 42 (28.1%) were further merged with codes with similar meaning but different wording. Finally, 116 indicators, 16 subcomponents, and 4 components were synthesized.

Step 3: Interpreting the Framework

The nature of and associations among the components, subcomponents, and indicators were further assessed by the author group. Each indicator was identified as either a facilitator or a barrier for data collection. Data and themes that had been extracted from expert elicitation were constantly compared between hospitals and CDCs with varying data collection processes and contexts and between different data collection roles played by different experts. The data analysis outputs were triangulated and discussed within the team until a consensus was reached (Figure 1).

XSL•FO
RenderX

Figure 1. The validation process for the 4D framework.



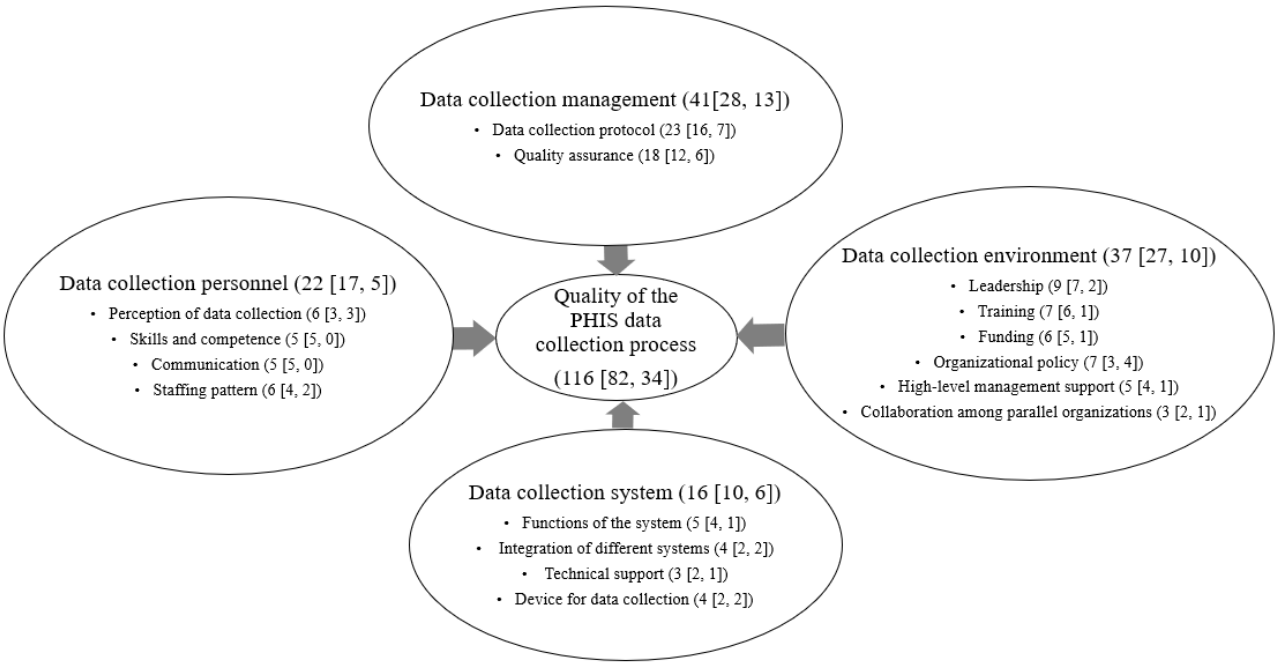
Results

Overview

The 4 dimensions of the 4D framework were verified as data collection management, data collection environment, data collection personnel, and data collection system. Three new subcomponents were added: organizational policy, high-level management support, and collaboration among parallel organizations. A total of 16 subcomponents were validated and grouped into the appropriate 4D components. They were measured by 116 indicators, including 82 facilitators and 34 barriers (Figure 2).

XSL•FO
RenderX

Figure 2. Composition of the 4D framework. Parenthesis: (number of indicators [number of facilitators, number of barriers]). PHIS: public health information system.



The next section presents the results using the 4D components to tabulate and elaborate the evidence that supports the subcomponents and indicators of the validated 4D framework situated in the CRIMS.

Data Collection Management

Data collection management includes 2 essential subcomponents: *data collection protocol* and *quality assurance*. Of the 302 interview codes, 88 (29.1%) supported 59 original indicators of data collection management. The remaining 14 were deleted because of a lack of evidence support. Furthermore, 5 new ones emerged from the interview codes. After merging 23 supported original indicators to amalgamate similar meaning with different wording, 41 indicators, including 28 facilitators and 13 barriers, were finalized for measuring the data collection management (Multimedia Appendix 3).

Data Collection Protocol

A total of 56 interview codes were related to the subcomponent *data collection protocol*. They validated 23 indicators, including 16 facilitators and 7 barriers, and fell under the subdimension of data collection form and data collection methods.

Six practitioners (C302, C303, C201, C101, C106, and A101) suggested that the data collection protocol should be aim-focused, operable, scientific, rational, and feasible for frontline data collectors. It should contain comprehensive, step-by-step guidance for the entire process of data collection (A101 and C201). The involvement of frontline data collectors in the development of a data collection protocol was an optimal practice (C302).

A total of 16 practitioners (C101, C102, C103, C104, C105, C106, C107, C201, C203, C303, C304, A101, H302, H305, H101, and H202) stressed that a data collection form needs to be clear, readable, comprehensive, and unambiguous. One of the practitioners mentioned:

> *It [design of the form] needs to be rational to make data collection convenient and simple, and provides comprehensive data, should reduce data collectors burden and reduce unnecessary effort.* [C102]

To ensure that the questions about data collection are articulated in a scientific, rational, and operable manner, 3 CDC practitioners (C201, C202, and C107) recommended the following: (1) to solicit a question, one can ask questions from different angles; (2) the number of questions should be suitable and controlled within the allotted data collection time; (3) the wording of the questions, including options for the multiple-choice questions, must be accurate, direct, understandable, and answerable; and (4) questions should be bound within ethical considerations and should not cause harm to respondents, particularly in places where it is challenging to find confidential and private space for question elicitation.

The data collection methods should be well developed, uniform, applicable, and implementable for data collectors (C302 and C301). A method is considered optimal for data collection if the task is integrated into routine data flow in a health care facility.

Quality Assurance

Overall, 32 interview codes were related to the subcomponent *quality assurance* and validated a total of 18 indicators, including 12 facilitators and 6 barriers. Three topics were elicited: the criteria of quality assurance, the constituency of quality assurance, and the implementation of conduct of quality assurance.

The criteria of quality assurance were consistent with the requirements of data quality, that is, accuracy, completeness, and timeliness. Therefore, quality assurance is “able to

XSL•FO
RenderX

thoroughly, quickly and accurately assess data accuracy, completeness and timeliness" (C203).

Clinicians believed that data quality audits were useful in motivating data collectors because their managers may provide extra funding to incentivize this activity. H201 explained the advantage of a data audit:

> *On one side, it is useful to provide further verification guidance to our routine work, and correct deficiencies inevitable in operational procedures because we are new to this job. On the other side, if they could brief the findings to our manager, it would be even better...For example, if my workplace was equipped with the needed amenities, then it will be easy and convenient. It does not necessarily need further monetary injection.* [H201]

Two health administrators (A401 and A301) who held a position at the national and provincial administration used the CRIMS data regularly for decision making. They relied on the quality of the data quality assessment conducted by all levels of CDCs. "The professionals will ensure the quality of data collection process" (A401), whereas "it is impossible to verify the situation (of data) in the front line" (A301).

A401 expressed his concern about the deliberate falsification of data, especially the *soft* data. Soft data means that its data quality is difficult to assess even with field verification, such as data from high-risk population intervention, follow-up, and health education. *Hard* data are more likely to be *true*, for instance, the methadone treatment data documented on the site, and thus, *hard* data are less prone to errors:

> *It does have difference in level of data accuracy. Some data are relatively accurate, such as the data about methadone treatment because they were recorded when the patient took the medicine; that possibly would not be falsified, right? However, intervention data, the "relatively soft data", are hard to verify in the office. If you do not make an on-site verification, it is hard to control the recording of them.* [A401]

Data Collection Environment

The data collection environment includes 3 original subcomponents (leadership, training, and funding) and 3 newly added ones (organizational policy, high-level management support, and collaboration among parallel organizations). Of the 302 extracted codes, 86 (28.5%) were about the data collection environment, with 32 relating to the 3 new subcomponents. A total of 27 new indicators emerged from these interview codes. Of the 20 original indicators, 5 were deleted because of a lack of evidence and another 5 were merged further for a similar reason. A total of 37 indicators, including 27 facilitators and 10 barriers, were finalized to measure the data collection environment (Multimedia Appendix 3).

Leadership

Of the 28 practitioners, 26 (93%) agreed that leadership is a subcomponent of the quality of the CRIMS data collection process, ranking first among all subcomponents. Twenty-four codes were related to *leadership*. A total of 9 indicators, including 7 facilitators and 2 barriers, were validated for measuring leadership.

Concerning qualification and role, leaders should be role models with professionalism (C103, C105, C106, C203, and H304). They are **"**able to ensure the procedures to be executed up to standards, ensure the implementation of requirements and protocols of data collection, analysis, and use, and thus ensure data quality" (C203). To initiate a new task data collection, the leaders should have a clear roadmap for assigning and executing the task (A101). Leaders should have strong organizational capabilities to push it forward (H301, H303, and C106). Therefore, leaders do not necessarily have to do everything by themselves but should be familiar with the task requirements (C103, C104, and C105). They should have the power to issue policies, clarify and assign duties and tasks, and provide financial support (C302 and C104).

The extent to which a leader attaches importance to data collection determines the quality of this task. "People follow the example of their superiors" (H304 and H305). Clinicians (H202, H203, H301, and H306) mentioned that a significant indicator of adequate notice by a leader in charge is the frequency of his or her attending the meetings or the supervision and inspection events organized by the CDC.

From the practitioners' perspective, a good leader could "lead us well, ensuring those willing to do to have the opportunity to do, and turn those reluctant to do into willing to do; this is good leadership" (H304). The management recognition of the contribution of data collection personnel to data quality is an important motivation factor for data collectors (C102 and H305). It could be in the format of "oral praise to recognize and formal acknowledgement beyond financial incentives" (C102). As commented by H305 and H306, "our leaders all think highly of this job (data collection). Otherwise, the staff would not care." "Data collection personnel need to be respected, trusted, acknowledged, and complemented by leaders" (H304, C104, and C302).

Training

Of the 28 practitioners, 27 (93%) agreed that training is a subcomponent of the quality of the CRIMS data collection process, ranking second among all components. A total of 19 interview codes about training generated and validated 7 indicators, including 6 facilitators and 1 barrier focusing on the objective of training, and the methods to deliver it and evaluate it.

The goal of training is to equip data collectors with qualified work competence and professionalism (C102, C103, C104, C105, C106, and C302):

> *The training objective is to equip the data collectors with work competence, with good work professionalism, such as work abilities and skills.* [C302]

> *I think training is more related to communication of [data collection] skills. Firstly, we must be familiar with the survey, then we will explore how we get good*

XSL•FO
RenderX

> *data. Learning skills is an objective to be reached via training.* [C104]

Therefore, training needs to focus on the standardization and uniformity of the data collection process. These include objective, methods, and time frames for data collection (C203, C103, C104, C106, and H304). The trainees should understand the definition of data to be collected, have good knowledge about all procedures for data collection, and adhere to the standardization.

Interactive training between trainers and trainees has been suggested (C103). During training, trainers should address work issues and help trainees learn what to do and how to do it (C103 and C105). Trainers should not just talk and go and be disinterested in whether the trainees understand or not. Trainers who were welcomed by trainees were those who quickly responded to trainees' questions (C105) and those providing empirical field practice examples in the training session. C103 suggested "if the trainers give more empirical examples for the training, the trainees may obtain a better understanding."

Data collection personnel, especially the newly recruited staff, need training after recruitment and refresh their knowledge every year about what and how to do. On-the-job training, hospital webpage training, and exams have been used in health facilities (H101 and H306). Building up a training network that installs materials and sources under the circumstance of high staff turnover is recommended (C106).

Given that the training results might vary among trainers, a training assessment was recommended, including selecting trainers, training methods, and training contents. C103 claimed that the higher the level of a training organization, such as international organizations and high-level CDCs, the better the training quality.

Funding

Although the subcomponent funding was not included in the item weight table, 10 relevant codes emerged from the interview transcripts and generated 3 new indicators. Three original indicators remained. They gave rise to 5 facilitators and 1 barrier to measuring the subcomponent funding.

From the CDC professionals' perspective, funding should support purchasing data collection devices such as computers, printers, and even vehicles (C301, C103, and C104). Funding should provide compensation, such as gifts for health clients to seek their cooperation (C103). Otherwise, "without funding support, without policy, and without competent personnel, data quality may be problematic, or even a fake product made up in office" (C103). Continuous funding support for previously funded projects is needed to avoid the situation of "when the Demonstration Project finished, funding decreased significantly" (C104).

From the hospital data registrars' perspective, HIV/AIDS work does not bring in profit, an activity that does not support the profit goal of the hospital (H301, H202, and H203):

> *HIV/AIDS prevention activities do not bring in profit, the doctors in the hospital should be committed and have spirit of dedication. However, in market economy, hospital needs profit, and is focused on pursuit of economic cost effectiveness.* [H301]

> *Without funding support, clinicians think they are volunteers for HIV/AIDS data collection. Therefore, sometimes, they are unwilling to do this job.* [H203]

Therefore, given that "funding support can provoke work" (H201) and "the cost of management and treatment can be reimbursed" (H203), "funding support for data collectors must be fully implemented" (H202 and H203). The health administrator (A101) had already recognized this need and promised to further push this agenda:

> *In another aspect, it might be related to boosting work morale to encourage them [data collection personnel] by increasing funding support. For example, they may get some subsidies for the work they are doing or have done. Currently we do have some funding. The performance-based salary system is inflexible. It may be problematic to link their income with their performance. This shortage, maybe, is what we need to tackle, for example, from the perspective of national management. We should be able to do, but not much has been done yet. They should get a better income. This aspect is what we should do.* [A101]

Organizational Policy

Organizational policy is a new subcomponent. Of the 28 practitioners, 23 (82%) agreed to place it in the component of the data collection environment. A total of 13 codes were related to the organizational policy and generated 7 indicators, 3 facilitators and 4 barriers. These indicators were primarily concerned with what organizational policy is desirable for HIV/AIDS data collection.

The organizational policy was critical to ensure the execution of the data collection activities (C104 and H101). "If they attach importance to the job, you will work easily; if they don't, your work is a challenge" (H306). It was regarded as more important than the actual process of data collection because the latter was under the control of the data collector (H303). The organizational policy should "support recognition and reflection of the real situation and encourage analysis of existing issues, a proactive adaptation of scientific findings generated from analysis of high-quality data" (C203).

Desired organizational policies of the CRIMS data collection process included (1) ensuring sufficient funding, staffing and material support, for example, "as long as the workload is increased, more staff is assigned" (H101); (2) embodying good management and coordination; (3) having a built-in reward and bonus scheme (C301 and C202) to "motivate people to work well" (H303).

Indicators of a poor organizational policy relating to data collection included the following aspects: (1) data collection was set up as a part-time job, (2) narrow workspace insufficient for data collection (H302), (3) increased workloads without adequate payment (H201), and (4) the culture of "eating big-pot-rice" (C106).

XSL•FO
RenderX

High-Level Management Support

High-level management support was another newly added subcomponent that 79% (22/28) of the practitioners agreed to. A total of 19 interview codes generated 5 indicators, including 4 facilitators and 1 barrier, to measure this subcomponent of high-level management support.

High-level management support provides assurance (C201); assistance for training; responsibility for policymaking (H305); and being scientific, specific, and rigorous for decision making (C104 and C106). It enforces an appropriate reward and punishment mechanism (H303). High-level management support also means funding support and making essential data collection tools such as vehicles available (C103).

Conversely, high-level management support should "not impose excessive administrative pressure on data collectors because it may compromise data integrity and accuracy in data collection. The management should not affect and intervene in the data collection process and the data. Otherwise, it may cause a negative consequence of manipulating results" (C203). In practice, the policy had a significant impact on the data quality (C104 and C302). The health administrator (A301 from the provincial health bureau) had a different viewpoint: "currently, as for the HIV/AIDS epidemic data collection, indeed there is no intervening in our work, basically it (data) is reliable, no concealment."

High-level management support was recognized as "a strong power that can veto by just a couple of words" (A101 from the county health bureau). The more the emphasis on data quality placed by upper management, the more time would be invested by data collectors toward data quality and vice versa (C102). "No site auditing, no proper work" (H303). However, the more the layers between the high-level management and the frontline data collection organization, the more difficult it is for the organization to execute the data collection process (C104).

The CDCs were considered by H201, a clinician in a secondary hospital, as "supportive" high-level management; they were expected to provide hospitals with support and advocacy. The CDCs were also expected to be of help and to understand "why, what and how" about data collection. H301, a data registrar in a tertiary hospital, suggested that the local CDC should "clarify the work-flow in hospital and do not just require us doing this and doing that without distinction."

Collaboration Among Parallel Organizations

Collaboration among parallel organizations was a third newly added subcomponent, with 82% (23/28) practitioners agreeing. A total of 14 interview codes were related to this subcomponent, which may contribute to HIV/AIDS data collection, for example, through hospitals and CDCs. Furthermore, 3 indicators, including 2 facilitators and 1 barrier, were added to the 4D framework to measure collaboration among parallel organizations.

It was found that sometimes the quality of the data collected by the collaborating organizations may have inferior quality to those collected by the CDCs, if they are without staff in charge. Therefore, if data to be collected were provided by a collaborating organization, C403 suggested a coordinating HIV/AIDS committee would contribute to high-quality data collection. He stressed,

> *"If the parallel organizations with dependency in data do not have a right attitude toward data collection, or conduct data collection in a reckless manner, then the data to be collected would be worse (in quality) and useless.* [C403]

A public health professional (C203) working at a city-level CDC stated that the parallel organization "should not use vicious competition and negative approaches to intervene with public health data collection and organizations. They should cooperate, coordinate and facilitate."

Data Collection Personnel

The component *data collection personnel* included 4 essential subcomponents: *perception of data collection*, *skills and competence*, *communication*, and *staffing patterns*. Of the 302 interview codes, 77 (25.5%) supported 20 of the 29 original indicators of the data collection personnel in the preliminary framework. Six new indicators emerged, and 4 were merged further. There were 22 indicators, including 17 facilitators and 5 barriers for measuring data collection personnel.

Perception of Data Collection

All 28 practitioners agreed that data collectors' perception of data collection is an important subcomponent determining the quality of the CRIMS data collection process. Of the 6 original indicators about the perception of data collection, 4 were supported by the interview transcripts, 1 was deleted because of a lack of evidence, and the other was merged with 2 newly added indicators. Six indicators, including 3 facilitators and 3 barriers, were finalized to measure the perception of data collection.

From some practitioners' perspectives (C102, C103, H306, H202, and H203), the CRIMS data collection process would not be as complicated if the data collection personnel were aware of its importance, which would also lead to better data quality. As H203 said*:*

> *It is a matter of how serious they (doctors) are definitely. Because this (data collection) is a very simple and easy job. If you pay attention to it, you can do it well.* [H203]

H202 and H203, 2 public health data registrars working in a secondary hospital, agreed that the priority given by clinicians and managers in the hospital could significantly improve the quality of the data collection process and thus data quality:

> *It is an issue of whether the doctors and management value it (data collection). If the management values data collection, doctors will also value the activity.* [H203]

It was suggested that the *perception of data collection* should not only be measured by receptibility to data collection but also by 2 new indicators, including commitment of the data collection personnel to data collection and their attitude to integrity (C103, C201, C203, C302, and H203). The manifestation of *good* attitude may be "data were consistent between the paper-based

XSL•FO
RenderX

and the electronic records of the CRIMS" (C103). The fabrication of data or negligence is often caused by poor attitudes rather than incompetence or lack of training for data collection. Burnout demotivates data collection personnel to treat the job as their job responsibility. C106, a public health professional with 8 years of work experience at a county CDC for HIV/AIDS prevention and control, suggested that burnout may appear after working on the same job for a long period. "Now nobody values much about this job, so not many are willing to do it, including me" (C106).

Skills and Competence

All 28 practitioners agreed that data collection skills and work competence were important for data collection personnel. Five indicators, all facilitators, were recommended for measuring subcomponent skills and competence.

This subcomponent was a *must-have* capability for frontline data collectors (C202), which is more important than the data collector's education level (C201, C102, and C103):

> *If they [with high education degree] do not have adequate work experience, if they do not have work skills, they cannot find the solution to the problem.* [C201]

Besides the skills for data quality check, the subcomponent skills and competence includes an accurate understanding of the objective of data collection, contextual knowledge, and the definition of data items (C102, C103, C106, and H102). Data collection personnel should be able to make a rational judgment about the reliability of a data source and ensure data accuracy and completeness (C302, C202, C203, C104, C105, C106, A101, and H302). Communication, organization, coordination, and writing skills were also desired skillsets recommended by practitioners for a competence-based framework (A101, C302, C201, C102, H302, and H305).

The data collection personnel should be professional and receive training in data collection. Interns were not considered qualified for data collection and reporting. H302, a clinician from a tertiary hospital, suggested that work competence means being mature and experienced, which is not what an apprentice is up to. H301 and H306 reported that the interns in tertiary hospitals were asked to fill in the data collection forms for busy clinicians.

Communication

Although communication was not listed in the item weight table, a lack of good communication among data collectors, as described in the preliminary framework with 5 facilitative indicators, was verified by the practitioners, particularly those who need to directly interact with health clients in routine work (H302, H305, H201, and C106).

H201, an HIV/AIDS clinical specialist, felt embarrassed in detecting transmission routes through conversations with AIDS patients. She thought that transmission routes were a private issue, especially for young men. If the data to be collected do not affect treatment, then data quality can be compromised in the interest of preserving the privacy and dignity of patients:

> *All in all, it (knowing whichever transmission route) does not affect treatment. Through conversation with them, I feel that these patients are worried about we, doctors, are discriminating against them. This is the major concern. So, collecting this type of data (transmission route) is neglected in my job.* [H201]

C106, a county CDC professional, felt that it was difficult to communicate with AIDS patients during follow-up:

> *Sometimes, I do not even know how to communicate with them. Like meeting someone new, I am not sure what kind of psychological characteristics the person has. Basically, I feel them difficult to deal with. I do not even know how to talk to them. Sometimes it is fine; this feeling has always been there.* [C106]

She also felt that she was not getting adequate support from her colleagues in a routine job:

> *Having been in this job so long, it is embarrassing to ask others certain problems you encounter. You can only formulate solutions by yourself. You find it difficult to ask others. Better do it yourself.* [C106]

Staffing Pattern

Although the staffing pattern was not in the item weight table, it was mentioned by 11 practitioners. A total of 18 interview codes supported 6 of the 7 original indicators, including 4 facilitators and 2 barriers.

Practitioners frequently mentioned a lack of an adequate number of competent public health professionals:

There are only two staff members assigned to work at the front line of HIV/AIDS control by the Department of HIV/AIDS. These two staff members have to collect all data, and they are under enormous pressure; this indicates the staffing level is inadequate. [C103]

> *I feel the most challenging is staffing level. Lots of work needs people to do. It does not mean there is no staff to do the work but almost everyone has several parallel lines of work happening at a given time. Like us, old employees, all part time regarding data collection.* [C107]

In C107's workplace, employment of contractors was a major approach to fill the vacancy, but it was not favored by local public health professionals because of high turnover. The professionals even refused to train the contractors because they were worried that their efforts would be wasted if the contractors quit the job soon after the training was completed.

Experienced staff and female staff were considered (by C302, C201, C101, C106, and H302) to be the optimal personnel for collecting quality data, rather than young practitioners, because of their experience in interacting with and establishing rapport with AIDS patients. Four practitioners (C302, C105, C106, and H305) suggested that education level, training, experience, personality, and value could affect work competency and, thus, the quality of data collection.

The health administrator from the national Ministry of Health (A401) suggested a need to increase the recruitment of frontline data collectors to cope with the increased workload in HIV/AIDS prevention and control.

XSL•FO
RenderX

Data Collection System

The component data collection system includes 4 subcomponents: functions of the system, integration of different information systems, technical support, and devices for data collection. A total of 51 codes for this component were identified, which supported 67% (18/27) of the original indicators about the data collection system in the preliminary framework and generated 9 new indicators. After comparison, 11 original indicators were further merged. A total of 16 indicators, including 10 facilitators and 6 barriers, were developed to measure the component data collection system (Multimedia Appendix 3).

Functions of the Data Collection System

A total of 17 interview codes were related to the subcomponent *functions of the data collection system.* They supported 8 original indicators of this subcomponent. Two new indicators emerged, and 5 original indicators were merged. A total of 5 indicators, including 3 facilitators and 1 barrier, were finalized to measure the subcomponent functions of the data collection system.

The practitioners agreed that the functions of the CRIMS should facilitate the visualization of routinely collected data. The CRIMS system should be humane, convenient, and error-free for data collection. For example, the system should remind data collectors wherever logic errors or incompleteness appear in data entry. In H304's words, "Machine can do something for human beings."

In recognition of the effect of *smart chart* and drop-down menus, some practitioners (C202, H302, and H305) suggested that the CRIMS should provide a user-friendly interface, allowing clinicians to add descriptive free text data; visualize data; and search by keywords, such as symptoms of a disease. The system should have convenient or automatic functions, such as iPhone's one-click for all end users, and should eliminate tedious extra work. The hospital practitioners were not satisfied with the CRIMS menu allowing limited details. It was inconvenient and difficult for H303 to add additional text data:

Some definitions are too narrow. For example, loss for follow-up could have a variety of reasons in reality, but we could not enter these data. Another example is the patient background. They may have lots of opportunistic infections without clinical symptoms; however,there are not enough options provided by the system to capture all situations. [H303]

An information system without adequate functions may impair data quality. C301 spent nearly 15 minutes, one-third of her interview time, to elaborate on this topic according to her work experience. Ascertain definitions of data items in the system were not in accordance with those of the data collectors, which may lead to missing data or inaccurate data collection.

Integration of Different Information Systems

The interview transcripts supported 4 of the 7 original indicators that discussed the integration of different information systems. Four indicators, 2 facilitators and 2 barriers, were clarified for measuring this subcomponent.

Although the item "integration of different information systems" was not in the weighting table, the negative effects caused by the lack of integration of data across information systems were emphasized by practitioners from hospitals (H302, H303, H304, H305, and H201). Hospital information systems are internal systems that do not connect to external systems via the internet. Access to the CRIMS was only available on a few authorized computers in hospitals via internet connectivity. Clinicians could not read any information from the CRIMS beyond their hospital. Repetition in reporting often happened, causing *a wasted job* that could lead to clinicians' reporting cards being "thrown into a rubbish bin" (H305). Therefore, it is a common sentiment that appropriate integration of the CRIMS with hospital information systems is needed.

In addition, 6 practitioners raised the importance of comprehensive data storage in the CRIMS information system (A101, A401, C106, H302, H303, H305). They suggested the system should include all work functions and topics, and cover all geographic regions from village, county, city to the province and national levels. From the national health administrator's perspective (A401), the CRIMS should be such a system:

> *From the perspective of a specific case reporting system, I think, it is a very comprehensive system; maybe no other disease reporting system can be as comprehensive as it is. The AIDS (CRIMS) should be the most comprehensive one.* [A401]

Technical Support

A total of 12 codes identified from the transcripts discussed technical support. Two new indicators emerged and supported the original indicators in the preliminary framework. Three indicators, 2 facilitators and 1 barrier, were finalized for measuring this subcomponent.

Practitioners (C302, C202, C104, C105, and H101) stated that insufficient technical support could inhibit the quality of the CRIMS data collection process. They emphasized that technical support should also be available for data entry. Technical support differed from training. It should be available before and during data collection. Practitioners from the county CDCs (C104 and C105) recommended that technical support for data entry should include a multimedia-supported electronic network that stores frequently asked questions, allows end users to share experiences, and provides help to use the system. It should offer access to higher level support such as that from national institutions. Technical support must be comprehensive, problem-focused, and formal.

Technical support became exceptionally critical for a data collection task assigned by high-level authority without training. Given that data collection tasks were often directly deployed by the high-level authority through issuing an official notification (C104), A101 believed that a competent team leader could play a role in offering technical support.

Devices for Data Collection

Of the 5 original indicators about devices for data collection, 4 were supported and the other was merged. A new indicator emerged and was added. Four indicators, 2 facilitators and 2

XSL•FO
RenderX

barriers, were finalized to measure the subcomponent. The compatibility of the devices used for data collection with the CRIMS data collection system was a major concern.

The practitioners suggested that data collection devices should be of good quality (C106, C104, C102, and H101), reliable, fast, and fit for surfing the internet and should neither crash nor break down (C302 and C304). Prompts, such as "the system is under maintenance" (C102), were not welcomed by practitioners. They expected that the devices could help them perform their data work even at the peak time of data entry. It should be free from traffic jams (H306, H305, C203, and C102).

Regarding data backup and security, the CRIMS has specific policies requesting the duration of data storage and the frequency of data backup to mitigate the risk of data loss (H306).

Discussion

Principal Findings

This study used the expert elicitation research method to verify a preliminary 4D component framework for measuring the quality of the PHIS data collection process in the context of the Chinese HIV/AIDS information management system, the CRIMS. The 28 public health data management experts for the CRIMS, with varied work experience and roles in their organizations, provided insightful inputs to issues related to the quality of the data collection process. They agreed with the 4 main components derived from the literature [12]. They ranked and commented on the importance of the original subcomponents based on their perceptions of the CRIMS data collection process. The 302 codes identified from the interview transcripts supported 75.2% (112/149) of the original indicators. These results provided the basis for a validated 4D component framework that fits well with the preliminary framework.

The validated 4D component framework was an improvement on the preliminary version. New items were identified in the expert elicitation process and added to the subcomponents of the data collection environment. These were organizational policy, high-level management support, and collaboration with parallel organizations. A total of 46 new indicators were generated and integrated into the framework, showing a wide range of characteristics elicited from the specific research context.

The original indicator statements were further simplified, merged, or deleted based on the 3 data analysis scenarios. The number of indicators in the framework finally decreased from 149 to 116.

There were changes within the framework in the proportions of the indicators for the 4 main components. The proportion of the indicators of the data collection environment increased from 13.4% to 31.9%, that for data collection management decreased from 49.0% to 35.3%, and that for the data collection system decreased from 18.1% to 13.8%. There was no change in the proportion for data collection personnel (19%). The factors that affect the quality of the data collection process are multifaceted from the practitioners' perspective.

Lessons Learned

The detailed feedback from the participants provided deep insights into many issues related to the quality of the data collection process and matters that require ongoing negotiation and development to improve it.

Under data collection management, the methods and protocols with the third ranking among all subcomponents need to be well developed, uniform, and implemented by data collection personnel. Responses on quality assurance emphasized the importance and challenges of this area. In some cases, data collection protocols and quality assurance procedures were developed and issued by high-level management in public health, but frontline personnel were not involved. This might make the data collection protocol and methods not operable or unfeasible in the field. Strategies to improve data collection management should include the involvement of frontline public health data collectors, especially those in hospitals, in the design phase of data collection protocols and quality assurance procedures [29].

A friendly data collection environment is an indispensable component of a high-quality public health data collection process. Participants ranked leadership and training as the two most important items for this component. This is consistent with and corroborated by the International Standard Organization's recommendation that the top management should "demonstrate leadership and commitment with respect to the quality management system" [30].

Various identified organizational issues complemented the subcomponent spectrum for the data collection environment. This included avoidance of data collection intruding unduly on health facilities' operations, such as routine health services in hospitals. This also included the adequacy of communication between different organizations, such as health administration and hospitals, CDCs and hospitals, and between data collection staff and their superiors. Financial and logistical support for the data collection process appeared to be a major issue, as is the case for health care organizations in many countries [29,31-33]. If the level of support is inadequate or not suitably administered, data quality will deteriorate.

On the data collection personnel component, all practitioners agreed on the importance of work attitude, competence, and data audit skills. There was some variation in opinion regarding the difficulty of the data collection process. The priority placed by the management in a hospital that performs the data collection process can significantly affect performance [34]. The fabrication of data or negligence indicated a poor attitude, requiring action by managers and supervisors. *Burnout* exhibited by staff might appear after long-term work in data collection and would require remediation, especially in hospitals [34].

Work competence was considered as a *must-have* capability for frontline data collectors. In addition to data quality audit skills, there should be an understanding of the objective of data collection and the definition of data items. Increasing the number of competent staff would, in principle, help to improve the data collection quality, although a practical difficulty has been the high turnover of recruited contract staff following training [32].

XSL•FO
RenderX

The fourth component, the data collection system, is an area that is influenced by the continuing changes in the performance and availability of ICTs [35]. Functions in the system should facilitate the visualization of routinely collected data. The system should be humane for those who operate it and be convenient and error-free for data collection [35]. An inappropriate function in the system may impair data quality. For example, if the definition of data items in the system does not reflect the reality of the work undertaken, the results will be unconvincing.

As identified in the preliminary 4D component framework, insufficient technical support inhibits the quality of the PHIS data collection process [12]. Additional features suggested by participants included storage of frequently asked questions and shared experiences, help for staff using the system, and access to higher level support such as that from national institutions. They also saw a need for the integration of the data collection system with other information systems because disconnection may result in repetitive reporting and inappropriate use of resources [5,33].

A study contribution is that, for the first time, we confirmed that the 4D components provide a picture of the structure and operation of the HIV/AIDS data collection process in China. The findings suggested that the Chinese HIV/AIDS information management practice provided an effective validation case and enriched the field of the quality of the PHIS data collection process. Three new subcomponents—organizational policy, high-level management support, and collaboration among parallel organizations—were considered to influence the quality of China's public health data collection process. This provided evidence to clarify the effect of the data collection environment on the quality of the CRIMS data collection process. The 4D framework also advocates the involvement of relevant stakeholders in data quality management. This provides an example to suggest the potential of using this framework for root cause analysis to investigate and identify the *real* factors behind poor data quality.

Although this study provides useful inputs to management decisions within the CRIMS and to negotiations with other parties on resources and requirements, it is reasonable to believe that the framework is also applicable to other settings, such as emerging infectious disease surveillance [36], general health care, education, and criminal justice.

Comparison With Prior Work

The context of this investigation was the Chinese HIV/AIDS program. However, many of the issues identified in the 2 sources of validation, the CRIMS and China, are also echoed in other health care systems. Inadequate staff training for data collection and limited support were also reported in birth registration in the United States [37] and in antiretroviral treatment for HIV infection in South Africa [38]. Poor communication across the health care sector and between providers was found in Aboriginal cardiac rehabilitation in Australia [39]. A lack of data linkage and sharing in electronic immunization data collection systems was described in Canada [40]. Job fatigue was found in general practitioners in European countries [34]. Regarding the transferability or generalizability embedded in the findings, this validation study has achieved design validity via a well-organized and executed research process [26].

As there are few extant public health frameworks focusing on the quality of the data collection process, there is a genuine contribution that this research has made to fill a critical gap on this topic. The successful abstraction of the 4D framework components, subcomponents, and indicator statements demonstrates the need for qualitative research in a problem domain without known measurement methods. Therefore, this study has taken the right method and approach given the novelty of the research topic, despite its importance in ensuring public health data quality.

Limitations

A potential limitation of this study is that a relatively small sample of experts participated in the interview, which may be small for statistical probabilistic generalizability. The control strategy was to use the theoretical sampling method, including all levels and types of participating organizations, personnel roles, and experts in the CRIMS system. This eventually brought data saturation for qualitative inquiry and provided comprehensive views of the HIV/AIDS data collection process in China. Given that the purpose of this study was to use a qualitative method to validate a preliminary conceptual framework, we have achieved our aim.

Although the number of indicators was reduced from 149 to 116, these indicators need further item reduction for ease of use in large-scale public health settings. This can be achieved by conducting quantitative questionnaire surveys with public health data management personnel at all levels. This will improve the validity of the 4D component framework and allow the reduction of measurement items to a manageable level.

Conclusions

This qualitative study validated 4D components for the quality of the PHIS data collection process in the context of the Chinese HIV/AIDS information management system, the CRIMS. The findings demonstrate that data collection management, data collection environment, data collection personnel, and data collection system are key components that determine the quality of the Chinese HIV/AIDS data collection process. The 4D component framework was further modified into a new pool containing 16 subcomponents and 116 indicators. They can be further tested and judged by practitioners and researchers in future public health data quality assessment studies.

Acknowledgments

This research was conducted with the support of an Australian Government Research Training Program Scholarship.

XSL•FO
RenderX

Conflicts of Interest

A potential conflict of interest may arise as HC was personally known to some of the experts because of her familiarity with the Chinese public health professionals, health administrators, and medical institutions in the study area.

Multimedia Appendix 1
Original 4D components of the quality of the public health information system data collection process [<xref ref-type="bibr" rid="58ref7">7</xref>, <xref ref-type="bibr" rid="58ref12">12</xref>-<xref ref-type="bibr" rid="58ref15">15</xref>].
[DOCX File , 16 KB - jmir_v23i5e17240_app1.docx]

Multimedia Appendix 2
Agreement with an item being a component or subcomponent and its importance rank for the quality of the Comprehensive Response Information Management System data collection process (N=28).
[DOCX File , 16 KB - jmir_v23i5e17240_app2.docx]

Multimedia Appendix 3
Indicators, including facilitators and barriers, in each subdimension of the 4 dimensions of the quality framework of the data collection process for public health information systems.
[DOCX File , 28 KB - jmir_v23i5e17240_app3.docx]

References

1. World Health Organization. World health statistics 2017: monitoring health for the SDGs, Sustainable Development Goals. Geneva Switzerland: World Health Organization; 2017.
2. World Health Organization. Monitoring the Building Blocks of Health System: A Handbook of Indicators and Their Measurement Strategies. Geneva, Switzerland: World Health Organization; 2010.
3. Lin V, Smith J, Fawkes S, Robinson P, Chaplin S. Public Health Practice in Australia: The Organised Effort. Crows Nest, N.S.W: Allen & Unwin; 2007.
4. World Health Organization. Data quality review: a toolkit for facility data quality assessment. Module 1. Framework and metrics. Geneva, Switzerland: World Health Organization; 2017:5-12.
5. Mao Y, Wu Z, Poundstone K, Wang C, Qin Q, Ma Y, et al. Development of a unified web-based national HIV/AIDS information system in China. International Journal of Epidemiology 2010 Nov 26;39(Supplement 2):ii79-ii89. [doi: 10.1093/ije/dyq213]
6. Davis J, Morgans A, Burgess S. Information management in the Australian aged care setting. Health Inf Manag 2017 Jan;46(1):3-14. [doi: 10.1177/1833358316639434] [Medline: 27105477]
7. Heywood A, Boone D. Guidelines for data management standards in routine health information systems. United States Agency for International Development (USAID)/MEASURE Evaluation. 2015. URL: https://www.measureevaluation.org/resources/publications/ms-15-99 [accessed 2021-04-20]
8. Sheikhtaheri A, Nahvijou A, Sedighi Z, Hadji M, Golmahi M, Roshandel G, et al. Development of a tool for comprehensive evaluation of population-based cancer registries. International Journal of Medical Informatics 2018 Sep;117:26-32. [doi: 10.1016/j.ijmedinf.2018.06.006]
9. Dufour É, Duhoux A, Contandriopoulos D. Reliability of a Canadian database for primary care nursing services' clinical and administrative data. International Journal of Medical Informatics 2018 Sep;117:1-5. [doi: 10.1016/j.ijmedinf.2018.05.007]
10. Liu Y, Wu Z, Mao Y, Rou K, Wang L, Zhang F. Quantitatively monitoring AIDS policy implementation in China. International Journal of Epidemiology 2010 Nov 26;39(Supplement 2):ii90-ii96. [doi: 10.1093/ije/dyq214]
11. Karr AF, Sanil AP, Banks DL. Data quality: A statistical perspective. Statistical Methodology 2006 Apr;3(2):137-173. [doi: 10.1016/j.stamet.2005.08.005]
12. Chen H, Yu P, Hailey D, Cui T. Identification of the essential components of quality in the data collection process for public health information systems. Health Informatics J 2019 May 29:146045821984862. [doi: 10.1177/1460458219848622]
13. Australian IOHW. The nKPI data collection: data quality issues working paper 2012-2014. Canberra: AIHW; 2015:2014.
14. Canadian IFHI. The CIHI data quality framework. Ottawa, Ont: CIHI; 2009.
15. O'Carroll P, Yasnoff W, Ward M, Ripp L, Martin E. Public health informaticsinformation systems. Hannah K, Ball MJ. editors. New York, NY: Springer Science & Business Media Inc; 2003.
16. Knol AB, Slottje P, van der Sluijs JP, Lebret E. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environ Health 2010 Apr 26;9(1):19. [doi: 10.1186/1476-069x-9-19]
17. Chen H, Yu P, Wang N. Do we have the reliable data? An exploration of data quality for AIDS information system in China. Studies in Health Technology and Informatics.: IOS Press; 2013. URL: https://pubmed.ncbi.nlm.nih.gov/23920816/
18. Mokkink L, Terwee C, Patrick D, Alonso J. Stratford PaulW, Knol DirkL. COSMIN checklist manual. 2012. URL: https://docplayer.net/29633833-Cosmin-checklist-manual.html [accessed 2019-06-18]

XSL•FO
RenderX

19. Grigore B, Peters J, Hyde C, Stein K. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. BMC Med Inform Decis Mak 2017 Sep 4;17(1):131. [doi: 10.1186/s12911-017-0527-0]
20. Hitimana R, Lindholm L, Mogren I, Krantz G, Nzayirambaho M, Sengoma JS, et al. Incremental cost and health gains of the 2016 WHO antenatal care recommendations for Rwanda: results from expert elicitation. Health Res Policy Sys 2019 Apr 5;17(1):36. [doi: 10.1186/s12961-019-0439-9]
21. Cooke R, Shrader-Frechette K. Experts in Uncertainty : Opinion and Subjective Probability in Science. Cary, United States: Oxford University Press, Incorporated; 1991:9780195362374.
22. Wu Z, Wang Y, Detels R, McGoogan J. HIV/AIDS in China Epidemiology, Prevention and Treatment [eBook]. Singaporepringer Nature Singapore Pte Ltd. 2020. URL: https://doi.org/10.1007/978-981-13-8518-6 [accessed 2020-07-16]
23. Li L, Cao XB, Luo W, Wang CH, Wu ZY, Rou KM. [Quality of network direct reporting of information about intervention service in population at high risk for HIV infection in China, 2013-2014]. Zhonghua Liu Xing Bing Xue Za Zhi 2017 Sep 10;38(9):1197-1200. [doi: 10.3760/cma.j.issn.0254-6450.2017.09.011] [Medline: 28910931]
24. Li L, Luo W, Xu J, Rou K. Assessment of data quality of intervention in population at high risk for HIV/AIDS in China, 2014-2018. Chin J Epidemiol 2020;41(1):85-89. [doi: 10.3760/cma.j.issn.0254-6450.2020.01.016]
25. Chen H, Yu P, Hailey D, Cui T. Data Quality of the Chinese National AIDS Information System: A Critical Review. Stud Health Technol Inform 2017;245:1352. [Medline: 29295431]
26. Venkatesh V, Brown SA, Bala H. Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. MISQ 2013 Jan 1;37(1):21-54. [doi: 10.25300/misq/2013/37.1.02]
27. Zhang Y, Yu P, Shen J. The benefits of introducing electronic health records in residential aged care facilities: A multiple case study. International Journal of Medical Informatics 2012 Oct;81(10):690-704. [doi: 10.1016/j.ijmedinf.2012.05.013]
28. Pope C, Ziebland S, Mays N. Analysing qualitative data. British Med J 2000;320(320):6804-6806. [doi: 10.1136/bmj.320.7227.114] [Medline: dsjsr.2518]
29. Andersson A, Hallberg N, Timpka T. A model for interpreting work and information management in process-oriented healthcare organisations. Int J Med Inform 2003;72(1):47-56. [Medline: 14644306]
30. International Organization for Standardization. ISO 9001 Quality management systems - requirements. Geneva, Switzeland: International Organization for Standardization (ISO); 2015:2015.
31. Hahn D, Wanjala P, Marx M. Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics. Global Health Action 2013 Aug 29;6(1):21424. [doi: 10.3402/gha.v6i0.21424]
32. Yip JLY, Bright T, Ford S, Mathenge W, Faal H. Process evaluation of a National Primary Eye Care Programme in Rwanda. BMC Health Serv Res 2018 Dec 7;18(1):950. [doi: 10.1186/s12913-018-3718-1]
33. Lopez AD, Setel PW. Better health intelligence: a new era for civil registration and vital statistics? BMC Med 2015 Apr 7;13(1):73. [doi: 10.1186/s12916-015-0333-4] [Medline: 25885782]
34. Le Floch B, Bastiaens H, Le Reste JY, Lingner H, Hoffman R, Czachowski S, et al. Which positive factors give general practitioners job satisfaction and make general practice a rewarding career? A European multicentric qualitative research by the European general practice research network. BMC Fam Pract 2019 Aug 9;20(1):96. [doi: 10.1186/s12875-019-0985-9]
35. Yu P, Zhang Y, Gong Y, Zhang J. Unintended adverse consequences of introducing electronic health records in residential aged care homes. International Journal of Medical Informatics 2013 Sep;82(9):772-788. [doi: 10.1016/j.ijmedinf.2013.05.008]
36. Cao Z, Tang F, Chen C, Zhang C, Guo Y, Lin R, et al. Impact of Systematic Factors on the Outbreak Outcomes of the Novel COVID-19 Disease in China: Factor Analysis Study. J Med Internet Res 2020 Nov 11;22(11):e23853. [doi: 10.2196/23853]
37. Melnik TA, Guldal CG, Schoen LD, Alicandro J, Henfield P. Barriers in Accurate and Complete Birth Registration in New York State. Matern Child Health J 2015 Feb 5;19(9):1943-1948. [doi: 10.1007/s10995-015-1711-4]
38. Kaposhi BM, Mqoqi N, Schopflocher D. Evaluation of antiretroviral treatment programme monitoring in Eastern Cape, South Africa. Health Policy and Planning 2014;30(5):547-554. [doi: 10.1093/heapol/czu028]
39. DiGiacomo M, Davidson P, Taylor K, Smith J, Dimer L, Ali M. Health information system linkage and coordination are critical for increasing access to secondary prevention in Aboriginal health: a qualitative study. Quality in Primary Care 2010;18(1):17-26.
40. Heidebrecht CL, Kwong JC, Finkelstein M, Quan SD, Pereira JA, Quach S, et al. Electronic immunization data collection systems: application of an evaluation framework. BMC Med Inform Decis Mak 2014;14:5 [FREE Full text] [doi: 10.1186/1472-6947-14-5] [Medline: 24423014]

Abbreviations

China CDC: Chinese Center for Disease Control and Prevention
CRIMS: Comprehensive Response Information Management System
ICT: information and communication technology
PHIS: public health information system

XSL•FO
RenderX

Edited by G Eysenbach, R Kukafka; submitted 28.11.19; peer-reviewed by T Wu, G Rao, N Liu, I Mircheva, R Poluru; comments to author 24.08.20; revised version received 06.11.20; accepted 11.03.21; published 10.05.21.

Please cite as:
Chen H, Yu P, Hailey D, Cui T
Validation of 4D Components for Measuring Quality of the Public Health Data Collection Process: Elicitation Study
J Med Internet Res 2021;23(5):e17240
URL: https://www.jmir.org/2021/5/e17240
doi:10.2196/17240
PMID:33970112

©Hong Chen, Ping Yu, David Hailey, Tingru Cui. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.05.2021. 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

Feasibility, Acceptability, and Efficacy of Virtual Reality Training for Older Adults and People With Disabilities: Single-Arm Pre-Post Study

Pui Hing Chau[1], BSocSc, PhD; Yan Yan Jojo Kwok[1], BN (Hons), MPH, PhD; Mee Kie Maggie Chan[1], RN, BN, MMSc, DHSc; Ka Yu Daniel Kwan[1], BSc; Kam Lun Wong[2], BSW, MSocSc; Ying Ho Tang[2], RSW, BAPAM, MSW; Kan Lung Peter Chau[2], PDOT, MCNS; Sheung Wa Matthew Lau[2], MScHC; Yan Yan Yannex Yiu[2], BSc (Hons), MSc; Mei Yan Fanny Kwong[2], BSW; Wai Ting Thomas Lai[3], HBSc, MALCS; Mun Kit Leung[2], RSW, BSW, MSocSc

[1]School of Nursing, The University of Hong Kong, Hong Kong, China

[2]Tung Wah Group of Hospitals, Hong Kong, China

[3]Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China

Corresponding Author:
Pui Hing Chau, BSocSc, PhD
School of Nursing
The University of Hong Kong
4/F, William MW Mong Block
21 Sassoon Road, Pokfulam
Hong Kong
China
Phone: 852 3917 6626
Email: phchau@graduate.hku.hk

Abstract

Background: Unlike most virtual reality (VR) training programs that are targeted at homogenous populations, a set of VR games for rehabilitation purposes targeted at a heterogeneous group of users was developed. The VR games covered physical training, cognitive training (classification and reality orientation), community-living skills training, and relaxing scenery experiences. Special considerations for local older adults and people with disabilities were made in terms of hardware choice and software design.

Objective: This study aimed to evaluate the feasibility, acceptance, and efficacy of VR training among users with varying abilities.

Methods: A single-arm pretest-posttest evaluation study was conducted. The participants of the evaluation study were encouraged to undergo 30-minute VR training three times a week for 6 weeks. The 30-minute session consisted of 10 minutes of upper-limb motion games, 10 minutes of lower-limb motion games, and 10 minutes of cognitive games/community-living skills training/relaxing scenery experiences, as appropriate. On completion of each session, usage statistics were documented via the built-in VR software, whereas feedback on the experience of the VR games and adverse events was collected via self-reports and staff observations. Feasibility was reflected by usage statistics, and acceptance was reflected by positive feedback. In addition, health outcomes, including upper-limb dexterity, functional mobility, cognitive function, and happiness, were assessed at baseline, as well as 6 weeks and 3 months after baseline. The primary outcomes were upper-limb dexterity and acceptance of playing VR games.

Results: A total of 135 participants with a mean age of 62.7 years (SD 21.5) were recruited from May 2019 to January 2020, and 124 (91.9%) completed at least one follow-up. Additionally, 76.3% (103/135) of the participants could attend at least 70% of the proposed 18 sessions, and 72.5% (1382/1906) of the sessions had a training time of at least 20 minutes. Linear mixed effect models showed statistically significant effects in terms of upper-limb dexterity (small effect) and cognitive function (moderate effect). Among the 135 participants, 88 provided positive comments. Additionally, 10.4% (14/135) reported mild discomfort, such as dizziness, and none reported severe discomfort.

Conclusions: A set of VR training games for rehabilitation could be applied to users with heterogeneous abilities. Our VR games were acceptable to local older adults and those with different disabilities. Benefits in upper-limb dexterity and cognitive

XSL•FO
RenderX

function were observed despite partial compliance to the training protocol. Service providers could refer to our experiences when developing VR training systems for their clients.

(J Med Internet Res 2021;23(5):e27640) doi:10.2196/27640

KEYWORDS

virtual reality; rehabilitation; older adults, people with disabilities; evaluation

Introduction

Virtual reality (VR) has been widely used in many areas, including education, aviation, medicine, and entertainment. For control and navigation in VR games, users are required to have relatively high cognitive and physical levels. In this regard, researchers and service providers are developing and refining various VR games to extend their applications to older adults and people with disabilities or diseases, who usually have longer response times and cannot handle controllers with buttons because of their weak fine-motor skills. With successful modifications, VR training provides these users the opportunity to engage in motivational training with many repetitions, salient stimuli, and challenging tasks [1]. Examples of VR training, particularly those with a rehabilitation purpose, are well documented in different populations, such as adults with stroke [2], community-dwelling older adults [3], people living with dementia and mild cognitive impairment [4], and children with cerebral palsy and Down syndrome [5].

VR games, particularly those related to living skills training and cognitive training, would be more appealing to users if they are based on local elements and culture that are familiar to the users. Hence, in our local setting (Hong Kong), there are various VR training programs targeting different populations, for example, people with intellectual disability [6,7], acquired brain injury or stroke [8-11], schizophrenia [12], dementia [13], high fall risk [14,15], and different geriatric conditions [16,17].

Despite many VR training programs being developed, they mainly target homogenous populations. There is a lack of an integrated VR training program that suits users with various types of conditions or disabilities. Therefore, there is a need to develop a universal VR training program in Hong Kong to serve a heterogeneous group of users comprising older adults and adults with different disabilities or conditions living in different settings. In 2019, one of the leading nongovernment organizations (NGOs) in Hong Kong took the initiative to develop a universal VR system to supplement the rehabilitation of older adults and people with disabilities, both in the residential care setting and the community setting, with a view to enrich the life experience of service clients and improve their training and treatment outcomes. The common rehabilitation goals of this heterogeneous population were upper-limb dexterity and cognitive status, since these are of utmost importance in maintaining activities of daily living for both institutional and community-dwelling populations. While mobility is also important, it was not one of the common goals, since some of the targeted users already lost this ability. Regarding such development, the efficacy of such VR training in improving health outcomes among users has to be scientifically evaluated. Therefore, the objective of this evaluation study was to investigate the feasibility, acceptance, and efficacy of VR training among older adults and people with different disabilities.

Methods

Study Design

This was a service evaluation study employing a single-arm pretest-posttest design. Although a randomized controlled trial or a quasiexperimental design was preferred, such designs would have to involve at least 145 participants per group in order to detect a small-to-medium effect. As the service units might not be able to support recruitment and intervention of such a large sample, these designs could not be adopted.

Participants

The target population included clients of the service units for rehabilitation or the service units for older adults of one NGO in Hong Kong. The inclusion criteria were service clients of the selected units, covering both residential care settings and community settings; age 18 years or above; ability to understand instructions in VR training; and presence of at least one of the following disabilities: functional impairment, mobility limitation, cognitive impairment, intellectual disability, and visual impairment. The exclusion criteria were conditions that restricted participation in VR training and severe discomfort with VR training.

The VR facility was expected to serve at least 130 clients. Depending on the functional status, cognitive status, and possible attrition among users, it was expected that at least half of the users (ie, 65 participants) would be eligible for outcome assessments. This would result in about 80% power and 5% level of significance to detect a small-to-medium effect size (Cohen d=0.35) in pre-post difference.

Interventions

A set of multipurpose immersive VR games was developed for VR training. These programs fit the language, content, and VR scenes specially designed for Hong Kong individuals, which enhanced usage experience. Various elements, including hardware, virtual setting, gaming operation, and game setting, were specially considered during program development to cater to users with different disabilities and functional levels, so as to ensure user friendliness and comfort for users (Table 1).

The components of the VR games included physical training, cognitive training (classification and reality orientation), community-living skills training, and relaxing scenery experiences (Table 2). To date, the types of VR training, as well as training intensity and frequency to achieve desired outcomes, have not been consistent. Benchmarking with other VR

XSL•FO
RenderX

programs for rehabilitation [18-21], the participants of the evaluation study were encouraged to undergo a 30-minute VR training three times a week for 6 weeks. The 30-minute session consisted of 10 minutes of upper-limb motion games, 10 minutes of lower-limb motion games, and 10 minutes of cognitive games/community-living skills training/relaxing scenery experiences, as appropriate.

Table 1. Some of the key considerations during program development to cater to the different conditions and disabilities of the users.

Element	Considerations
Hardware	Fully immersive simulation was adopted, with head-mounted display (HMD) devices and limb motion trackers used as the display method and control interface, respectively. Through the HMD, the images of virtual reality (VR) games were manifested in front of the users' eyes. It enabled users with low vision to see the VR scenery clearly. Moreover, the HMD device was highly portable and could be used with minimal space requirement, which fitted well for a densely populated place such as Hong Kong. Button-free limb motion trackers allowed the users to participate in VR games by simple limb movements without relying on fine-motor control.
Virtual setting	An approximately 180-degree interaction zone within the VR environments was adopted, where interactive objects or training tasks were set at arm's length or attached to users' virtual hands. Hence, users could play the games safely in a sitting posture without moving or bending forward. Agile body and head movements were excluded or minimized from the gaming design to alleviate cybersickness such as dizziness and nausea. The backdrop and characters of VR games were not too fancy or stimulating to cater to users with limited cognitive and visual perceptions.
Game operation	Users only needed to move the button-free trackers with their limbs to a target position and hold for about 2 seconds. Then, a preprogrammed action would be triggered. Moreover, some of the VR games allowed users to either stand up or sit down to play, whereas some permitted the use of one or more limbs to play selectively (eg, left, right, or both), whichever suits the user, such that even those with mutilation could enjoy VR games. Motion-detection sensitivity was tailor made to the capability of users so that even the smallest movement of their limbs and the slowest reaction time could be detected. Then, those with minimal mobility could also easily participate in VR games using button-free trackers.
Game setting	Most of the games were designed with three levels of difficulty so that users and/or staff could choose the most suitable one according to their abilities. All games have timely and encouraging feedback and cues. By providing immediate feedback, users would realize that what they do in the real world could trigger something in the VR games. In addition to feedback, cues were provided in the form of images, words, and sounds, which served as guides to help users understand what they should do in the VR games. Staff could also use their own cues as additional guidance. Only positive reinforcement was adopted in the game design, allowing users to freely explore the VR games without any sanction and penalty (eg, score deduction). It also provided a sense of achievement to users and encouraged them to continuously engage in VR training.

Table 2. Virtual reality games in the training session.

Component	Virtual reality (VR) games
Physical	The games involved the upper limbs (*handball*), lower limbs (*football*), or full-body motions (*gatekeeping*). Users repeatedly mobilized their limbs and used their muscles, which aimed to help slow down the deterioration of bodily functions.
Cognitive (classification skills)	Nostalgic elements were employed to appeal to the users. Three signature games/scenes from a local amusement park that operated from the 1940s to the 1990s were included. The games, namely *Elephant Feeding*, *Feather Duster Throwing*, and *Coin Tossing*, were familiar to most Hong Kong older adults. They could provide not only cognitive stimulation but also a reminiscence of the old days. Another game, named *Home Items Locating*, simulated a home setting and required users to identify the appropriate items in the corresponding rooms.
Cognitive (reality orientation)	In the game named *Hong Kong Footprint*, users were virtually brought to familiar local places, including Shek Kip Mei, Man Mo Temple, Wong Tai Sin Temple, and Chun Yeung Street. Each scene was played with a soundtrack that briefly introduced the place. Along the VR journey, a trained staff member discussed the place that the user was watching. The experience allowed older adults to revisit familiar places and widened the horizons of residential care clients. These aimed to elicit pleasure and satisfaction, improve communication, and enhance self-esteem and social skills.
Community- living skills	The game *Seven Must-Dos Before Leaving Home* aimed to promote users' awareness (especially older adults) of the seven safety measures that keep them and the living environment safe before leaving their homes. Another game named *MTR GO GO GO* aimed to help those with disabilities to develop skills in taking public transport and going to different places in Hong Kong by themselves. It simulated a Mass Transit Railway (MTR) station scene and aimed to familiarize users with travelling via the MTR under various conditions repeatedly and purposefully.
Relaxing scenery experiences	Four relaxing scenes, including *diving*, *river*, *starry sky*, and *grassland*, accompanied by relaxing music, were provided. The experience aimed to relieve users' physical tension, calm their agitated mood, and engage with them in an interactive virtual environment. In the virtual scenes, various interactive objects were merged with a real 3D scene. Users could interact with them and trigger a particular preprogrammed effect. Additionally, the scenes aimed to provide dramatic multisensory stimulation for some frail users who could not engage in real sceneries that require limb motions.

XSL•FO
RenderX

Procedure

Trained staff of the NGO identified and recruited potential participants to join the evaluation study. Upon recruitment, written informed consent was sought from the participants or their legal guardians. Eligibility screening and assessment were conducted by trained research assistants. Then, eligible participants engaged in 30-minute VR training three times a week for 6 consecutive weeks (giving a total of 18 sessions). Deviations were allowed according to the actual situation of the users. For example, if a week of training was missed due to sickness, the user was allowed to continue the program in the seventh week.

For each session, a trained staff member always accompanied the participant in the same room. The staff member was responsible to assist the user in putting on the head-mounted display (HMD) device and limb motion trackers (both hand and foot), to guide the user in playing the VR games, and to closely monitor the conditions of the user during game play to ensure safety. In between different training components, the participants could take off the HMD device and rest if needed. Any training component that was not applicable to the users would be skipped, for example, those with total loss of lower-limb function would not engage in lower-limb motion training. The participants could terminate the session whenever they wished. Adverse events were monitored and recorded as appropriate. After each session, the staff asked the participants for feedback on their experience of the VR games in that particular session.

Data Collection and Outcome Measures

Demographic and medical information, which included age, gender, living arrangement (community-dwelling vs institutional setting), education level, dependence status, mobility status, cognitive status, intellectual disability status, visual impairment status, medical history (eg, stroke and fracture), and medical conditions (eg, mental illness and autism), were extracted from the records of the residents/members at baseline.

Health outcomes, including upper-limb dexterity, functional mobility, cognitive status, and happiness, were assessed by the trained research assistants at baseline (T0), 6 weeks after baseline (T1), and 3 months after baseline (T2). Participants with conditions that prohibited them from performing certain assessments were excluded from the analysis for that outcome.

Upper-limb dexterity was measured by the standard Box and Block Test (BBT). The BBT has been shown to be a valid measure of dexterity in the older population [22]. The participants were asked to transfer blocks from one compartment to the other in 1 minute according to a standard procedure [23]. The number of blocks transferred to the second compartment is counted, and a larger number indicates better dexterity.

Functional mobility was measured by the Timed Up and Go Test (TUG). The participants were asked to get up from an armchair, walk 3 meters, turn back, and return to a seated position according to a standard protocol [24]. The time required to finish the task at each round was recorded. Three measurements were made. Shorter walking time indicated better balance. The best (fastest) time of the three TUG trials was used [25].

Cognitive status was measured using the Montreal Cognitive Assessment 5-Minutes (Hong Kong Version) (HK-MoCA 5-Min) protocol. The validated assessment test covered the following four domains: attention, executive functions/language, orientation, and memory [26,27]. The total score ranged from 0 to 30, with a higher score indicating better cognitive status. The alternate version was not used in subsequent assessments, as literature suggested that the results were unlikely to be affected if the same assessment tool was used over the follow-up period [28]. If the participants were incapable of answering the HK-MoCA 5-Min, they were examined using the Benton Temporal Orientation Test (BTO), which involves only questions about orientation to time (including identifying the year, month, day, day of the week, and time of the day) [29]. An error score was calculated according to the Benton Temporal Orientation Scale (BTOS), with a lower error score indicating better cognitive status [30].

Happiness was measured using a single question “for most of the time, do you feel…” [31], accompanied with a 11-point Likert-scale ruler or 5/3/2 facial expression pictures about happiness [32]. Following the pretest protocol suggested for the Personal Wellbeing Index–Intellectual Disability (Chinese-Cantonese), the intellectual ability of respondents was assessed to determine the response options [32]. The options (for respondents with descending order of ability) included an 11-point scale (0-10), a 5-point scale (0-4), a 3-point scale (0-2), and a 2-point scale (0-1). Then, the score was standardized into units of “percentage of scale maximum,” with a possible range of 0 to 100 and higher scores indicating higher levels of happiness [32].

Usage time of VR training was collected via the built-in VR software system to ensure data accuracy. At the end of each session, the trained staff who accompanied the participants in the VR training collected qualitative feedback on the training experience, using the question “Do you like or dislike the VR training? Why?” The participants could provide multiple feedback responses, both positive and negative, if they wished. Any adverse health effects (such as dizziness) induced after playing the VR games were also recorded through self-reports and staff observations. The staff accompanying the participants in the VR training closely observed any discomfort experienced by the participants, in addition to explicitly asking the participants about any discomfort.

Statistical Analysis

The primary outcomes were upper-limb dexterity and acceptance of VR training. The secondary outcomes were usage statistics, functional mobility, cognitive status, and happiness. SPSS version 25 (IBM Corp) was used for statistical analyses. A 5% significance level was adopted.

The participant characteristics were summarized using descriptive statistics. Characteristics of dropout participants were examined using a logistic regression model. VR training usage statistics in terms of the number of weeks engaged in VR training, number of total sessions of VR training attended, and duration of each VR training session were calculated and used to reflect the feasibility of the intervention. Linear mixed effect models were used to analyze the change in health outcomes,

XSL•FO
RenderX

adjusting for age, gender, education level, living arrangement, disability, and medical history/condition. Assessment time was used as the independent variable, and its fitted coefficients indicated the temporal change in health outcomes in T1 and T2 as compared with T0. Missing data at follow-up were not imputed as the linear mixed effect models could cater for missing data. The standardized effect size in terms of the Cohen *d* index was calculated using the adjusted mean difference divided by the standard deviation [33]. Cohen *d* index values of 0.2, 0.5, and 0.8 indicated small, moderate, and large effect sizes, respectively.

Content analysis was performed for the qualitative feedback on the experience of VR training. First, the presence of concepts related to positive and/or negative comments was determined. Then, the presence of keywords and related concepts, such as tiredness, equipment, excitement, and exercise, was counted. Lastly, the frequency and percentage were calculated accordingly. The acceptance level of VR training experience was reflected by the number of participants providing positive comments, with a larger proportion indicating a higher level of acceptance. As the participants could provide multiple feedback responses, both positive and negative, when we calculated the proportion of participants using the total number of participants as the denominator, the proportion could exceed 100%. Therefore, in order to investigate factors associated with a higher level of acceptance, participants had to be further classified into the following three mutually exclusive groups: (1) only or mostly positive comments, (2) only or mostly negative comments, and (3) no comments or equal positive/negative comments. The proportions of participants in each group were calculated. Multiple multinomial regression was conducted to identify the associated factors. Adverse health events were also summarized.

Ethics Approval

This study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (reference number: UW19-336).

Results

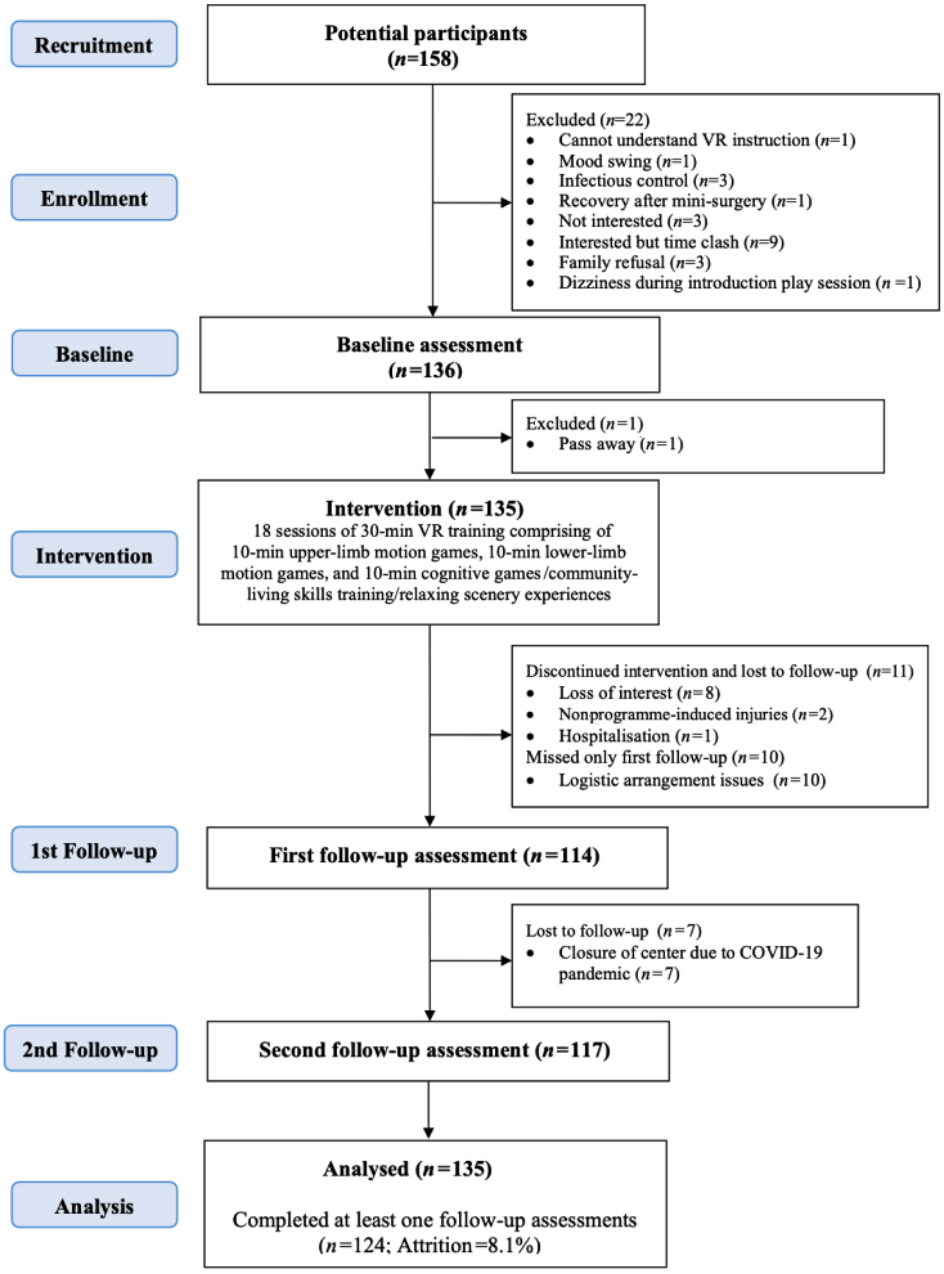
Characteristics of the Participants

A total of 158 participants were recruited from both institutional and community settings, including four homes for the severely disabled, eight homes for the intellectually disabled, six homes for the visually impaired, one old-age home, one community center for the intellectually disabled, and three elderly community centers. Of the 158 participants, 23 (14.6%) were not eligible, giving a total of 135 participants for the evaluation study (Figure 1). Table 3 summarizes the baseline characteristics of the 135 participants. The participants had a mean age of 62.7 years (SD 21.5). Additionally, 50.4% (68/135) of the participants were male, 76.3% (103/135) had an education level of primary school or below, or special educational needs, and 70.4% (95/135) were living in residential care settings. All participants had at least one type of disability, with 52.6% (71/135) having moderate-to-severe functional dependence, 78.5% (106/135) requiring walking aids or wheelchairs, 28.1% (38/135) having mild-to-severe cognitive impairment, 54.8% (74/135) having mild-to-severe intellectual disability, and 45.9% (62/135) having mild-to-severe visual impairment. Moreover, 9.6% (13/135) of the participants had a stroke history, 9.6% (13/135) had a fracture history, 13.3% (18/135) had mental illness, and 9.6% (13/135) had autism.

Additionally, 91.9% (124/135) of the eligible participants completed at least one follow-up. Some of them dropped out because of loss of interest (n=8), while a few of them did so because of nonprogram-induced injuries (n=2) and hospitalization (n=1). The logistic regression model showed that none of the participant characteristics was associated with dropout.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. VR: virtual reality.



XSL•FO
RenderX

Table 3. Baseline characteristics of the participants (N=135).

Characteristic	Value, n (%) or mean (SD)
Mean age (years)	62.7 (21.5)
Gender	
Male	68 (50.4)
Female	67 (49.6)
Educational level	
Primary or below, or special educational needs	103 (76.3)
Secondary or tertiary	32 (23.7)
Living arrangement	
Community dwelling	40 (29.6)
Residential care setting	95 (70.4)
Dependence level	
Independence or slightly dependence	64 (47.4)
Moderate-to-severe dependence	71 (52.6)
Mobility status	
Without auxiliary equipment	29 (21.5)
Require walking aids	34 (25.2)
Wheelchair bound	72 (53.3)
Cognitive status	
No cognitive impairment	97 (71.9)
Mild-to-severe cognitive impairment	38 (28.1)
Intellectual disability status	
No intellectual disability	61 (45.2)
Mild-to-severe intellectual disability	74 (54.8)
Visual status	
No visual impairment	73 (54.1)
Mild-to-severe visual impairment	62 (45.9)
History of stroke	
No	122 (90.4)
Yes	13 (9.6)
History of fracture	
No	122 (90.4)
Yes	13 (9.6)
Mental illness	
No	117 (86.7)
Yes	18 (13.3)
Autism	
No	122 (90.4)
Yes	13 (9.6)

Usage Statistics of VR Training

A total of 1906 VR training sessions were conducted during the study period. Table 4 shows the summary of the usage statistics of VR training. Ninety-one individuals participated in the VR training for 6 weeks consecutively and another 15 participated for at least 6 weeks intermittently, giving a total of 106 participants (106/135, 78.5%) that met the suggested usage

XSL•FO
RenderX

duration of 6 weeks. Over three quarters (103/135, 76.3%) of participants attended at least 13 sessions of VR training (ie, at least 70% of the proposed 18 sessions), and the majority (n=1382, 72.5%) of the 1906 sessions had a training time (excluding set-up time and rest time) of at least 20 minutes. The median total training time of the participants was 316.8 minutes. All of them participated in physical training games for a median of 275.0 minutes, and 116 of them participated in cognitive training (classification) games for a median of 33.8 minutes (Table 5). As for the difficulty level, 84.7% (1512/1786) of the *handball* game sessions, 88.7% (1295/1460) of the *football* game sessions, and 66.7% (909/1363) of the *classification* game sessions were at the easiest level (level 1) (Table 6). On the other hand, participants were able to play more advanced levels in *MTR GO GO GO*. Reasons for not complying with the usage protocol, including medical appointments or other engagements, limited resources, bad weather, technical problems, disability constraints, mood swings, sickness, and tiredness, were encountered in the scheduled sessions.

Table 4. Usage statistics of virtual reality training.

Usage statistics	Value, n (%)
Number of weeks of VR[a] training (n=135 participants)	
≥6 weeks consecutively	91 (67.4)
≥6 weeks intermittently	15 (11.1)
<6 weeks	29 (21.5)
Number of sessions of VR training (n=135 participants)	
≥18 sessions	28 (20.7)
13-17 sessions	75 (55.6)
<13 sessions	32 (23.7)
VR training time per session (n=1906 sessions)	
≥30 minutes	175 (9.2)
20-29 minutes	1,207 (63.3)
<20 minutes	524 (27.5)

[a]VR: virtual reality.

Table 5. Game time statistics according to game components.

Games	Number of participants	Time (minutes)					
		Mean	Median	Minimum	Maximum	SD	Range
All	135	289.3	316.8	6.0	595.0	129.6	6.0-595.0
Physical	135	244.0	275.0	4.0	390.0	108.6	4.0-390.0
Cognitive (classification)	116	43.0	33.8	2.0	177.8	35.0	2.0-177.8
Cognitive (reality orientation)	21	21.5	17.8	1.4	49.1	14.6	1.4-49.1
Community-living skills	18	16.8	15.1	3.3	42.2	12.4	3.3-42.2
Relaxing scenery experiences	27	13.7	10.5	2.6	50.0	12.1	2.6-49.9

Table 6. Difficulty levels of some virtual reality games played in the 1906 sessions.

Game	Sessions, n (%)			
	Level 1	Level 2	Level 3[a]	Mixed levels
Handball (n=1786 sessions)	1512 (84.7)	107 (6.0)	18 (1.0)	149 (8.3)
Football (n=1460 sessions)	1295 (88.7)	82 (5.6)	6 (0.4)	77 (5.3)
Classification (n=1363 sessions)	909 (66.7)	63 (4.6)	210 (15.4)	181 (13.3)
MTR GO GO GO (n=47 sessions)	9 (19.1)	22 (46.8)	15 (31.9)	1 (2.1)

[a]Including levels 3 and 4 in MTR GO GO GO.

XSL•FO
RenderX

Health Outcomes

According to their conditions, the participants received varying health outcome assessments at baseline, with the BBT (n=134 for the dominant hand and n=128 for the nondominant hand), TUG (n=104), HK-MoCA 5-Min (n=94), BTO (n=30), and happiness scale (n=130) (Table 7). Among the health outcomes, significant improvement over time was noted in upper-limb dexterity (P=.008 for the dominant hand and P=.043 for the nondominant hand) and cognitive function (P<.001 for overall performance) (Figure 2), whereas there was no significant change in functional mobility (P=.14), orientation to time (P=.72), and happiness (P=.34).

Table 7. Baseline outcome measures of the participants (N=135).

Outcome measure	Number of participants	Value, mean (SD)
BBT[a] (in blocks)		
Dominant hand	134	27.7 (14.1)
Nondominant hand	128	28.0 (14.7)
TUG[b] (in seconds)	104	15.4 (11.8)
HK-MoCA 5-Min[c] score		
Overall performance	94	13.7 (7.8)
Attention domain	94	2.3 (1.4)
Executive functions/language domain	94	3.9 (2.3)
Orientation domain	94	3.6 (2.1)
Memory domain	94	3.9 (3.2)
BTO[d] score (in error scores)	30	37.8 (25.9)
Happiness score	130	79.1 (30.0)

[a]BBT: Box and Block Test.

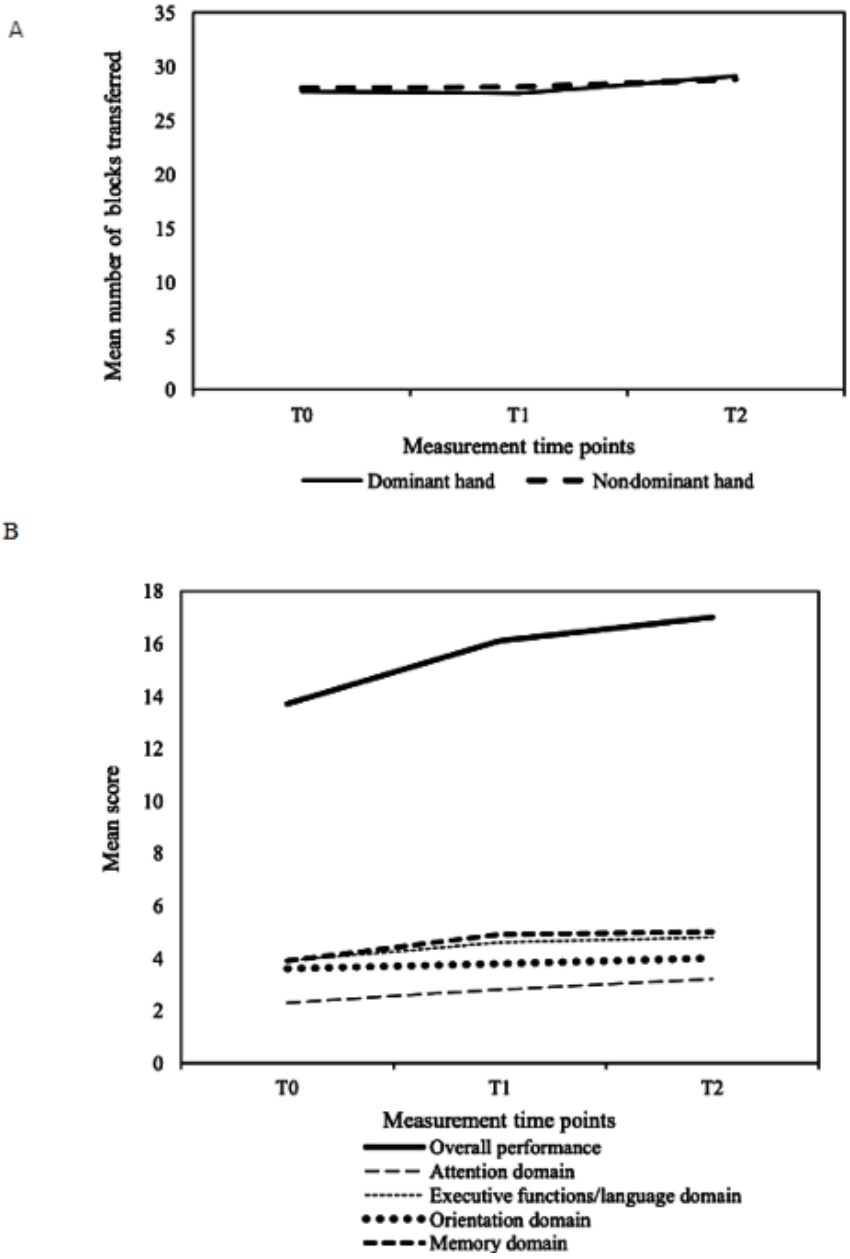
[b]TUG: Timed Up and Go Test.

[c]HK-MoCA 5-Min: Montreal Cognitive Assessment 5-Minute (Hong Kong Version).

[d]BTO: Benton Temporal Orientation Test.

XSL•FO
RenderX

Figure 2. Change in health outcomes over three measurement time points (baseline [T0], 6 weeks after baseline [T1], and 3 months after baseline [T2]). (A) Block and Block Test (BBT), a larger number of blocks transferred indicates better upper-limb dexterity; *P* values for time difference: .008 (dominant hand) and .043 (nondominant hand). (B) Montreal Cognitive Assessment 5-Minutes (Hong Kong Version) (HK-MoCA 5-Min), a higher score indicates better cognitive status; *P* values for time difference: <.001 (overall performance), <.001 (attention domain), <.001 (executive functions/language domain), .004 (orientation domain), and <.001 (memory domain).



Based on the linear mixed effect models (Table 8), the number of blocks transferred by the dominant hand significantly increased by 2.4 blocks (*P*=.007, Cohen *d*=0.17) and that by the non-dominant hand increased by 1.9 blocks (*P*=.04, Cohen *d*=0.13) when comparing T2 with T0. However, no significant differences in the number of blocks transferred were observed for both the dominant and nondominant hands when comparing T1 with T0.

In terms of the HK-MoCA 5-Min, the overall performance score significantly increased by 2.5 (*P*<.001, Cohen *d*=0.31) when comparing T1 with T0, and increased by 3.5 (*P*<.001, Cohen *d*=0.45) when comparing T2 with T0. Significant differences were also detected in T1 and T2 compared with T0 for all domains, except the orientation domain where an insignificant difference was reported for the comparison between T1 and T0. When comparing T2 with T0, the Cohen *d* of the domains ranged from 0.24 for orientation to 0.65 for attention.

XSL•FO
RenderX

Table 8. Change in health outcomes over the three assessment time points.

Outcome measures[a]	T1[b] vs T0[c]			T2[d] vs T0		
	Change[e] (95% CI)	Effect size[f]	*P* value	Change[e] (95% CI)	Effect size[f]	*P* value
BBT[g] (in blocks)						
Dominant hand	0.7 (−0.9 to 2.3)	0.05	.39	2.4 (0.9 to 4.0)	0.17	.002[h]
Nondominant hand	0.7 (−0.7 to 2.2)	0.05	.33	1.9 (0.4 to 3.3)	0.13	.01[h]
TUG[i] (in seconds)	−0.5 (−1.6 to 0.6)	−0.04	0.39	−1.1 (−2.2 to −0.007)	−0.09	.048[h]
HK-MoCA 5-Min[j] score						
Overall performance	2.5 (1.7 to 3.2)	0.31	<.001[h]	3.5 (2.8 to 4.3)	0.45	<.001[h]
Attention domain	0.5 (0.2 to 0.7)	0.33	.001[h]	0.9 (0.6 to 1.2)	0.65	<.001[h]
Executive functions/language domain	0.8 (0.4 to 1.1)	0.33	<.001[h]	1.0 (0.7 to 1.3)	0.43	<.001[h]
Orientation domain	0.3 (−0.005 to 0.6)	0.14	.054	0.5 (0.2 to 0.8)	0.24	.001[h]
Memory domain	1.0 (0.5 to 1.4)	0.30	<.001[h]	1.2 (0.7 to 1.6)	0.37	<.001[h]
BTO[k] score (in error scores)	2.6 (−7.4 to 12.6)	0.10	.61	−1.5 (−11.4 to 8.3)	−0.06	.76
Happiness score	−4.4 (−10.6 to 1.9)	−0.15	.17	−3.5 (−9.8 to 2.7)	−0.12	.26

[a]A better condition is represented by an increase in transferred blocks in the BBT, a decrease in time in the TUG, an increase in the HK-MoCA 5-Min score, a decrease in error scores in the BTO, and an increase in the happiness score.

[b]T0: baseline.

[c]T1: 6 weeks after baseline.

[d]T2: 3 months after baseline.

[e]The change in health outcomes over time were estimated using the linear mixed effect models, controlling for age, gender, education level, living arrangement, dependence level, mobility status, cognitive status, intellectual disability status, visual status, history of stroke, history of fracture, mental illness, and autism.

[f]Effect size is expressed in terms of the Cohen *d* index, where 0.2 indicates a small effect, 0.5 indicates a medium effect, and 0.8 indicates a large effect.

[g]BBT: Box and Block Test.

[h]Significant at <.05 level of significance.

[i]TUG: Timed Up and Go Test.

[j]HK-MoCA 5-Min: Montreal Cognitive Assessment 5-Minute (Hong Kong Version).

[k]BTO: Benton Temporal Orientation Test.

Feedback From Participants

Among the 135 participants, 88 (65.2%) provided positive comments and 63 (46.7%) provided negative comments (multiple comments allowed), whereas 30 (22.2%) did not provide any comment. Among the 88 participants who provided positive comments, 53 (60.2%) reported the VR experience as innovative, interactive, fun, or exciting, 49 (55.7%) found the experience motivating, and 30 (34.1%) perceived the VR games as exercises. Meanwhile, among the 63 participants who provided negative comments, 26 (41.3%) complained about the VR equipment (such as the HMD device), 25 (39.7%) felt bored, 24 (38.1%) felt physically tired, 17 (27.0%) described the experience as scary, tense, or worrying, and 16 (25.4%) reported the usage as complicated. Moreover, 38 (60.3%) of those who indicated negative comments expressed such comments during the first two game sessions.

In terms of enjoyment, most participants enjoyed the *elephant feeding* game (n=11), followed by the *handball* game (n=5), *football* game (n=5), *diving scenery experience* (n=5), *feather duster throwing* game (n=3), *home items locating* game (n=3), and *coin tossing* game (n=1). At the same time, in terms of dislike, most participants disliked the *elephant feeding* game (n=11), followed by the *diving scenery experience* (n=3) and *seven must-dos before leaving home* game (n=1).

The participants were further classified into the following three groups: those who provided only or mostly positive comments (n=61, 45.2%), those who provided only or mostly negative comments (n=34, 25.2%), and those who had no comments or provided equal positive/negative comments (n=40, 29.6%). Based on multiple multinomial logistic regression, home-dwelling participants (*P*<.001) and those without autism (*P*=.008) had a much higher chance of providing positive comments. On the other hand, none of the participant characteristics was associated with a higher chance of expressing negative comments.

Apart from the general feedback, reports on mild discomfort were obtained in 25 of the 1906 (1.3%) game sessions (14 of 135 participants, 10.4%). The most common mild discomfort

XSL•FO
RenderX

was dizziness (nine sessions from seven participants), followed by eyestrain (six sessions from three participants), hand/leg pain (three sessions from two participants), blurred vision (two sessions from one participant), eye redness (two sessions from one participant), hand tremors (two sessions from one participant), and cramps (one session from one participant). Over half (13/25, 52.0%) of these mild discomforts occurred in the first four sessions of the VR training. None of the participants reported severe discomfort. Moreover, staff reported technical problems in 63 of the 1906 (3.3%) game sessions.

Discussion

Overview

This evaluation study explored the feasibility and acceptability of a universal set of VR training games for users with different abilities and conditions. Additionally, the improvement in health outcomes over time was assessed. The findings from this study can inform the future development and application of VR training to heterogeneous users in different settings.

Feasibility of Providing VR Training to Users With Different Disabilities

Few VR games have been designed for the broad coverage of different populations. Our experience revealed the possibility that people with different disabilities could enjoy the same set of VR games. With different training components, clients with different disabilities could participate in suitable types of training. Nevertheless, manpower resources had to be invested in addition to the VR system, since close supervision of the users by trained staff was required.

With a rehabilitation goal, we proposed a training protocol of 3 days a week for 6 weeks, with a target of 18 sessions. Over 76% of the participants attended 13 or more sessions. Adherence to the proposed number of training sessions was comparable to that reported in the literature [34]. Flexibility for makeup training was allowed if participants were not able to attend certain sessions. Usage statistics showed that the majority of sessions could offer at least 20 minutes of game time. However, the game time for cognitive training/community-living skills training/relaxing scenery experiences was substantially less than that for physical training. One explanation could be the limit of the session time. These components were placed as the last training components of each session, and thus, they might have been terminated if time was running out. Scheduling was a critical element to the successful implementation of the program. To facilitate a 30-minute training time, the manpower, equipment, and venue had to be made available for at least 45 minutes to allow for the necessary time for setup, instructions, and breaks. Moreover, as reality orientation, community-living skills training, and relaxing scenery experiences were available for use at a later stage, usage statistics were substantially less than those for the classification game. Despite the users' inability to attend all 18 sessions or spend 30 minutes of training time per session, there was a significant improvement in cognitive function and upper-limb dexterity. Hence, we propose to retain the current usage protocol until a new one with reduced usage is tested.

In our experience, mild discomfort was reported by 14 of 135 participants (10.4%). In a local study examining the effects of VR cognitive stimulation activity among community-dwelling older adults (n=236), 1.4% of them reported severe discomfort regarding fatigue and eye strain, blurred vision, and dizziness after 20 to 25 minutes of VR exposure [17]. However, mild discomfort was not reported. In another study conducted by US researchers, it was reported that 12% of older people with Alzheimer disease/mild cognitive impairment and 19% of older people without such conditions dropped out from the VR game because of simulation sickness [35]. The proportion of participants reporting adverse outcomes among our users was comparable, if not lower, than the proportions reported in the literature. Therefore, participants with disabilities would not put themselves at a higher risk of adverse events. At the same time, some participants reported physical tiredness after VR training. While this study did not quantitatively investigate tiredness as an outcome, a meta-analysis suggested an insignificant difference in terms of the presence of tiredness after VR training as compared with participants engaging in traditional exercises [36].

Acceptance of VR Training Among Users at an Advanced Age or With Various Disabilities

Generally, most of the participants accepted the use of the VR training, as 65.2% of the participants indicated positive comments. Those living in the community and those without autism tended to provide more positive remarks. This might be because these participants were more expressive of themselves. On the other hand, the HMD device and limb motion trackers might have induced discomfort, which was the major reason for negative comments. To enhance their experience and satisfaction in terms of fit and comfort, hardware designers may develop lighter devices. To accommodate the special needs of clients with limitations in head and neck movements, sensors should also be further adjusted so that clients do not need to maintain an upright position throughout the sessions.

Some participants also perceived the VR games as complicated, but they could still complete the session under the supervision of trained staff. Although some users claimed the VR training was boring, there was no evidence showing that such negative remarks were related to the proposed VR training for 18 sessions. Indeed, those who made negative comments had such perceptions in the first two sessions. This might suggest the need for more briefings or orientations prior to the use of VR training. In the current program, the scoring regime had been modified to increase the users' sense of achievement. However, there was no personalized scoring regime to differentiate game difficulty levels for different disabilities over time. In future development, such features should be incorporated for motivating clients to continue training and promoting a higher sense of achievement where progression could be felt by users.

The VR games that involved physical activity training were well received by users. It was interesting to learn that the VR cognitive-training game featuring the elephant kept in a zoo in Hong Kong in the 70s received both favorable and unfavorable reviews from the users. Some loved it probably because it recalled memories of the old days, while some disliked it

XSL•FO
RenderX

probably because they were scared by the elephant's trumpet, which was close to that in real life. Future design of VR scenery should take a balanced approach among excitement, calmness, and boredom.

Efficacy of VR Training in Improving the Health Outcomes of Participants

Using a single-arm pretest-posttest design, the evaluation study showed significant improvement in upper-limb dexterity assessed by the BBT and cognitive function assessed by the HK-MoCA 5-Min. According to the literature, an increase of 2 points in the MoCA full version score was generally considered as a minimal clinically important difference (MCID) for older adults or stroke survivors, which was about 0.5 standard deviations as defined by the distribution-based approach [36-38]. However, the MCID in terms of the MoCA 5-Min version has not been discussed in the literature. Hence, we used a distribution-based approach of 0.5 standard deviations, and considered a Cohen *d* index of 0.5 as the MCID. In our evaluation study, the HK-MoCA 5-Min overall score increased by 3.5 units as compared with baseline. In terms of the standardized effect size, the HK-MoCA 5-Min score had a Cohen *d* index of 0.45, which aligned with this reference. As understanding and following instructions to play VR games involve cognitive processing, the improvement in cognitive function might be enhanced by playing all the games and not necessarily restricting playing to the sessions involving cognitive VR games.

As for the BBT, there was no well-established MCID. Literature suggested that a difference of 6 to 8 blocks is considered as the smallest real difference beyond measurement error among stroke survivors [39], which would translate to a Cohen *d* index of 0.4 to 0.6 based on the standard deviation estimated in our study. Since the effect size of the BBT achieved in our study was only 0.13 to 0.17, we had to be cautious in interpreting the clinical significance of such a change. For functional mobility, although there was a favorable trend in improvement, such improvement did not reach statistical significance. It appeared that training intensity, particularly when participants took a sitting position when playing the games, was not sufficient to achieve improvement in functional mobility.

Our results are consistent with recent reviews, which found that VR training attained small-to-moderate effects in terms of physical performance [40] and cognitive function [41]. Despite the significant improvement in the oriental domain of the HK-MoCA 5-Min scale, there was an insignificant change in the BTO score. Further research is suggested to examine the reasons. Nevertheless, we would be cautious about clinical significance in terms of the improvement in the BBT. Our insignificant results in terms of mobility are also consistent with the absence of an effect in terms of gait as reported by a review [40].

The evaluation study did not show any significant change in the happiness level. This is consistent with a recent systematic review of VR exercise training that reported insignificant effects on psychological outcomes such as calmness and enjoyment [40]. The review also suggested that immersive components of VR training and longer follow-up periods tended to be associated with smaller benefits, but statistical significance was not found. As VR games are designed to promote functional and cognitive ability, if happiness is to be enhanced, the content of VR games might need to be adjusted and a regular playing schedule might not need to be applied.

A supplementary analysis testing the interactions between different subgroups (dependence level, mobility status, cognitive status, intellectual disability status, visual status, stroke history, fracture history, mental illness, and autism) and time points was performed using linear mixed effect models. Insignificant interaction terms were found for all those explored, except a few, which might have occurred by chance. There was insufficient evidence for the difference in the change of outcomes between different subgroups, and future studies could revisit this with a more vigorous design.

Strengths and Limitations

The strengths of this evaluation study were its large sample size, adoption of validated scales as outcome measures, and long follow-up period. Qualitative feedback from the participants also informed the strengths and weaknesses of the VR training program. However, the study has some limitations. First, the lack of a control group might have limited the interpretation of the results that only efficacy instead of effectiveness could be investigated. Second, people with hearing impairment, those who were bed bound, and those with disability in turnaround were not included in the study, as the current system could not cater to their special needs, and the potential benefits or adverse effects in these users are unclear from this study. The uses of various cognitive training components, community-living skills training, and scenery experiences were not standardized, and some games were not available at the beginning of the program, resulting in vast variations in game time for the components. In addition, progression of the game difficulty level was not standardized in the usage protocol, and this should be considered in future studies. As qualitative feedback responses were collected at the end of each game session, investigation by game components could not be performed. A future study should continue to investigate the optimal modalities in the domains of physical, cognitive, and psychological training for people with different vulnerable health conditions, and a controlled trial would be preferred.

Conclusions

A set of VR training games for rehabilitation could be applied to individuals with heterogeneous abilities. Our VR games were acceptable to local older adults and those with different disabilities. Benefits in upper-limb dexterity and cognitive function were observed despite partial compliance to the training protocol. Service providers could refer to our experiences when developing VR training systems for their clients.

XSL•FO
RenderX

Acknowledgments

This study was supported by the Innovation and Technology Fund for Better Living of the Hong Kong SAR Government.

Conflicts of Interest

None declared.

References

1. 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]
2. 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]
3. Corregidor-Sánchez AI, Segura-Fragoso A, Criado-Álvarez JJ, Rodríguez-Hernández M, Mohedano-Moriano A, Polonio-López B. Effectiveness of virtual reality systems to improve the activities of daily life in older people. Int J Environ Res Public Health 2020 Aug 28;17(17):6283 [FREE Full text] [doi: 10.3390/ijerph17176283] [Medline: 32872313]
4. D'Cunha NM, Nguyen D, Naumovski N, McKune AJ, Kellett J, Georgousopoulou EN, et al. A mini-review of virtual reality-based interventions to promote well-being for people living with dementia and mild cognitive impairment. Gerontology 2019 May 20;65(4):430-440 [FREE Full text] [doi: 10.1159/000500040] [Medline: 31108489]
5. Lopes JBP, Duarte NAC, Lazzari RD, Oliveira CS. Virtual reality in the rehabilitation process for individuals with cerebral palsy and Down syndrome: A systematic review. J Bodyw Mov Ther 2020 Oct;24(4):479-483. [doi: 10.1016/j.jbmt.2018.06.006] [Medline: 33218550]
6. Tam SF, Man DWK, Chan YP, Sze PC, Wong CM. Evaluation of a Computer-Assisted, 2-D Virtual Reality System for Training People With Intellectual Disabilities on How to Shop. Rehabilitation Psychology 2005;50(3):285-291. [doi: 10.1037/0090-5550.50.3.285]
7. Wu JK, Suen WW, Ho TM, Yeung BK, Tam AS. Effectiveness of two dimensional virtual reality programme and computer-assisted instructional programme in training Mass Transit Railway (MTR) skills for persons with mental handicap: A pilot study. Hong Kong Journal of Occupational Therapy 2005;15(1):8-15. [doi: 10.1016/s1569-1861(09)70029-4]
8. Lam YS, Man DW, Tam SF, Weiss PL. Virtual reality training for stroke rehabilitation. NeuroRehabilitation 2006;21(3):245-253. [Medline: 17167194]
9. Yip BC, Man DW. Virtual reality (VR)-based community living skills training for people with acquired brain injury: A pilot study. Brain Inj 2009 Dec;23(13-14):1017-1026. [doi: 10.3109/02699050903379412] [Medline: 19891532]
10. Fong KN, Chow KY, Chan BC, Lam KC, Lee JC, Li TH, et al. Usability of a virtual reality environment simulating an automated teller machine for assessing and training persons with acquired brain injury. J Neuroeng Rehabil 2010 Apr 30;7:19 [FREE Full text] [doi: 10.1186/1743-0003-7-19] [Medline: 20429955]
11. Yip BC, Man DW. Virtual reality-based prospective memory training program for people with acquired brain injury. NeuroRehabilitation 2013;32(1):103-115. [doi: 10.3233/NRE-130827] [Medline: 23422463]
12. Tsang MM, Man DW. A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. Schizophr Res 2013 Mar;144(1-3):51-62. [doi: 10.1016/j.schres.2012.12.024] [Medline: 23356951]
13. Man DW, Chung JC, Lee GY. Evaluation of a virtual reality-based memory training programme for Hong Kong Chinese older adults with questionable dementia: a pilot study. Int J Geriatr Psychiatry 2012 May;27(5):513-520. [doi: 10.1002/gps.2746] [Medline: 21681818]
14. Fu AS, Gao KL, Tung AK, Tsang WW, Kwan MM. Effectiveness of Exergaming Training in Reducing Risk and Incidence of Falls in Frail Older Adults With a History of Falls. Arch Phys Med Rehabil 2015 Dec;96(12):2096-2102. [doi: 10.1016/j.apmr.2015.08.427] [Medline: 26360975]
15. Tsang WW, Fu AS. Virtual reality exercise to improve balance control in older adults at risk of falling. Hong Kong Med J 2016 Feb;22 Suppl 2:S19-S22 [FREE Full text] [Medline: 26908338]
16. Chan TC, Chan F, Shea YF, Lin OY, Luk JK, Chan FH. Interactive virtual reality Wii in geriatric day hospital: a study to assess its feasibility, acceptability and efficacy. Geriatr Gerontol Int 2012 Oct;12(4):714-721. [doi: 10.1111/j.1447-0594.2012.00848.x] [Medline: 22469232]
17. Chan JYC, Chan TK, Wong MPF, Cheung RSM, Yiu KKL, Tsoi KKF. Effects of virtual reality on moods in community older adults. A multicenter randomized controlled trial. Int J Geriatr Psychiatry 2020 Aug;35(8):926-933. [doi: 10.1002/gps.5314] [Medline: 32346896]
18. Bang YS, Son KH, Kim HJ. Effects of virtual reality training using Nintendo Wii and treadmill walking exercise on balance and walking for stroke patients. J Phys Ther Sci 2016 Nov;28(11):3112-3115 [FREE Full text] [doi: 10.1589/jpts.28.3112] [Medline: 27942130]
19. Cho GH, Hwangbo G, Shin HS. The effects of virtual reality-based balance training on balance of the elderly. J Phys Ther Sci 2014 Apr;26(4):615-617 [FREE Full text] [doi: 10.1589/jpts.26.615] [Medline: 24764645]

XSL•FO
RenderX

20. Kamińska MS, Miller A, Rotter I, Szylińska A, Grochans E. The effectiveness of virtual reality training in reducing the risk of falls among elderly people. Clin Interv Aging 2018;13:2329-2338 [FREE Full text] [doi: 10.2147/CIA.S183502] [Medline: 30532523]
21. Lee MM, Lee KJ, Song CH. Game-based virtual reality canoe paddling training to improve postural balance and upper extremity function: a preliminary randomized controlled study of 30 patients with subacute stroke. Med Sci Monit 2018 Apr 27;24:2590-2598 [FREE Full text] [doi: 10.12659/MSM.906451] [Medline: 29702630]
22. Desrosiers J, Bravo G, Hébert R, Dutil É, Mercier L. Validation of the Box and Block Test as a measure of dexterity of elderly people: reliability, validity, and norms studies. Arch Phys Med Rehabil 1994 Jul;75(7):751-755. [Medline: 8024419]
23. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. Am J Occup Ther 1985 Jun;39(6):386-391. [doi: 10.5014/ajot.39.6.386] [Medline: 3160243]
24. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Mar;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
25. Kristensen MT, Ekdahl C, Kehlet H, Bandholm T. How many trials are needed to achieve performance stability of the Timed Up & Go test in patients with hip fracture? Arch Phys Med Rehabil 2010 Jun;91(6):885-889. [doi: 10.1016/j.apmr.2010.01.021] [Medline: 20510979]
26. Yeung PY, Wong LL, Chan CC, Leung JL, Yung CY. A validation study of the Hong Kong version of Montreal Cognitive Assessment (HK-MoCA) in Chinese older adults in Hong Kong. Hong Kong Med J 2014 Dec;20(6):504-510 [FREE Full text] [doi: 10.12809/hkmj144219] [Medline: 25125421]
27. Wong A, Nyenhuis D, Black SE, Law LS, Lo ES, Kwan PW, et al. Montreal Cognitive Assessment 5-minute protocol is a brief, valid, reliable, and feasible cognitive screen for telephone administration. Stroke 2015 Apr;46(4):1059-1064 [FREE Full text] [doi: 10.1161/STROKEAHA.114.007253] [Medline: 25700290]
28. Lebedeva E, Huang M, Koski L. Comparison of alternate and original items on the Montreal Cognitive Assessment. Can Geriatr J 2016 Mar;19(1):15-18 [FREE Full text] [doi: 10.5770/cgj.19.216] [Medline: 27076861]
29. O'Keeffe E, Mukhtar O, O'Keeffe ST. Orientation to time as a guide to the presence and severity of cognitive impairment in older hospital patients. J Neurol Neurosurg Psychiatry 2011 May;82(5):500-504. [doi: 10.1136/jnnp.2010.214817] [Medline: 20852313]
30. Benton AL, Vanallen MW, Fogel ML. Temporal orientation in cerebral disesase. J Nerv Ment Dis 1964 Aug;139:110-119. [doi: 10.1097/00005053-196408000-00003] [Medline: 14206449]
31. McCarron M, Swinburne J, Burke E, McGlinchey E, Carroll R, McCallion P. Patterns of multimorbidity in an older population of persons with an intellectual disability: results from the intellectual disability supplement to the Irish longitudinal study on aging (IDS-TILDA). Res Dev Disabil 2013 Jan;34(1):521-527. [doi: 10.1016/j.ridd.2012.07.029] [Medline: 23085501]
32. Cummins RA, Lau AL. Personal Wellbeing Index - Intellectual Disability (Chinese- Cantonese). Australian Centre on Quality of Life. 2005. URL: http://www.acqol.com.au/uploads/pwi-id/pwi-id-chinese-cantonese.pdf [accessed 2019-05-01]
33. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
34. Miller K, Adair B, Pearce A, Said C, Ozanne E, Morris M. 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]
35. Davis R, Ohman J, Weisbeck C. Salient cues and wayfinding in Alzheimer's disease within a virtual senior residence. Environ Behav 2017 Nov;49(9):1038-1065 [FREE Full text] [doi: 10.1177/0013916516677341] [Medline: 29230067]
36. Wu CY, Hung SJ, Lin KC, Chen KH, Chen P, Tsay PK. Responsiveness, minimal clinically important difference, and validity of the MoCA in stroke rehabilitation. Occup Ther Int 2019;2019:2517658 [FREE Full text] [doi: 10.1155/2019/2517658] [Medline: 31097928]
37. Wong GKC, Mak JSY, Wong A, Zheng VZY, Poon WS, Abrigo J, et al. Minimum clinically important difference of Montreal Cognitive Assessment in aneurysmal subarachnoid hemorrhage patients. J Clin Neurosci 2017 Dec;46:41-44. [doi: 10.1016/j.jocn.2017.08.039] [Medline: 28887072]
38. Nasreddine ZS, Patel BB. Validation of Montreal Cognitive Assessment, MoCA, alternate French versions. Can J Neurol Sci 2016 Sep;43(5):665-671. [doi: 10.1017/cjn.2016.273] [Medline: 27670209]
39. Chen HM, Chen CC, Hsueh IP, Huang SL, Hsieh CL. Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. Neurorehabil Neural Repair 2009 Jun;23(5):435-440. [doi: 10.1177/1545968308331146] [Medline: 19261767]
40. Ng YL, Ma F, Ho FK, Ip P, Fu KW. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: A systematic review and meta-analysis of randomized controlled trials. Computers in Human Behavior 2019 Oct;99:278-291. [doi: 10.1016/j.chb.2019.05.026]
41. Kim O, Pang Y, Kim J. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]

XSL•FO
RenderX

Abbreviations

BBT: Box and Block Test
BTO: Benton Temporal Orientation Test
BTOS: Benton Temporal Orientation Scale
HK-MoCA 5-Min: Montreal Cognitive Assessment 5-Minutes (Hong Kong Version)
HMD: head-mounted display
MCID: minimal clinically important difference
NGO: nongovernment organizations
TUG: Timed Up and Go Test
VR: virtual reality

Edited by R Kukafka, G Eysenbach; submitted 05.02.21; peer-reviewed by YTD Cheung, S Kramer; comments to author 03.03.21; revised version received 15.03.21; accepted 03.04.21; published 04.05.21.

Please cite as:
Chau PH, Kwok YYJ, Chan MKM, Kwan KYD, Wong KL, Tang YH, Chau KLP, Lau SWM, Yiu YYY, Kwong MYF, Lai WTT, Leung MK
Feasibility, Acceptability, and Efficacy of Virtual Reality Training for Older Adults and People With Disabilities: Single-Arm Pre-Post Study
J Med Internet Res 2021;23(5):e27640
URL: https://www.jmir.org/2021/5/e27640
doi:10.2196/27640
PMID:33944795

©Pui Hing Chau, Yan Yan Jojo Kwok, Mee Kie Maggie Chan, Ka Yu Daniel Kwan, Kam Lun Wong, Ying Ho Tang, Kan Lung Peter Chau, Sheung Wa Matthew Lau, Yan Yan Yannex Yiu, Mei Yan Fanny Kwong, Wai Ting Thomas Lai, Mun Kit Leung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2021. 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

Effect of Cognitive Training in Fully Immersive Virtual Reality on Visuospatial Function and Frontal-Occipital Functional Connectivity in Predementia: Randomized Controlled Trial

Jae Myeong Kang[1,2], MD; Nambeom Kim[3], PhD; Sook Young Lee[1], BSc; Soo Kyun Woo[1], MD; Geumjin Park[1], MA; Byeong Kil Yeon[4], MD, PhD; Jung Woon Park[5], PhD; Jung-Hae Youn[6], PhD; Seung-Ho Ryu[7], MD, PhD; Jun-Young Lee[8], MD, PhD; Seong-Jin Cho[1], MD, PhD

[1]Department of Psychiatry, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

[2]Brain Health Center, Gil Medical Center, Incheon, Republic of Korea

[3]Biomedical Engineering Research Center, Gachon University, Incheon, Republic of Korea

[4]Department of Psychiatry, Gyeonggi Provincial Medical Center Suwon Hospital, Suwon, Republic of Korea

[5]Department of Game Engineering and IT Convergence Engineering, Graduate School of Gachon University, Seongnam, Republic of Korea

[6]Department of Counseling Psychology, Cha University, Seongnam, Republic of Korea

[7]Department of Psychiatry, School of Medicine, Konkuk University Medical Center, Konkuk University, Seoul, Republic of Korea

[8]Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Seong-Jin Cho, MD, PhD
Department of Psychiatry
Gil Medical Center
Gachon University College of Medicine
21, 774-gil
Namdong-daero, Namdong-gu
Incheon, 21565
Republic of Korea
Phone: 82 32 460 8420
Email: sjcho@gilhospital.com

Abstract

Background: Cognitive training can potentially prevent cognitive decline. However, the results of recent studies using semi-immersive virtual reality (VR)-assisted cognitive training are inconsistent.

Objective: We aimed to examine the hypothesis that cognitive training using fully immersive VR, which may facilitate visuospatial processes, could improve visuospatial functioning, comprehensive neuropsychological functioning, psychiatric symptoms, and functional connectivity in the visual brain network in predementia.

Methods: Participants over 60 years old with subjective cognitive decline or mild cognitive impairment from a memory clinic were randomly allocated to the VR (n=23) or the control (n=18) group. The VR group participants received multidomain and neuropsychologist-assisted cognitive training in a fully immersive VR environment twice a week for 1 month. The control group participants did not undergo any additional intervention except for their usual therapy such as pharmacotherapy. Participants of both groups were evaluated for cognitive function using face-to-face comprehensive neuropsychological tests, including the Rey-Osterrieth Complex Figure Test (RCFT) copy task; for psychiatric symptoms such as depression, apathy, affect, and quality of life; as well as resting-state functional magnetic resonance imaging (rsfMRI) at baseline and after training. Repeated-measures analysis of variance was used to compare the effect of cognitive training between groups. Seed-to-voxel–based analyses were used to identify the cognitive improvement–related functional connectivity in the visual network of the brain.

Results: After VR cognitive training, significant improvement was found in the total score ($F_{1,39}$=14.69, *P*=.001) and basic components score of the RCFT copy task ($F_{1,39}$=9.27, *P*=.005) compared with those of the control group. The VR group also showed improvements, albeit not significant, in naming ability ($F_{1,39}$=3.55, *P*=.07), verbal memory delayed recall ($F_{1,39}$=3.03, *P*=.09), and phonemic fluency ($F_{1,39}$=3.08, *P*=.09). Improvements in psychiatric symptoms such as apathy ($F_{1,39}$=7.02, *P*=.01), affect ($F_{1,39}$=14.40, *P*=.001 for positive affect; $F_{1,39}$=4.23, *P*=.047 for negative affect), and quality of life ($F_{1,39}$=4.49, *P*=.04) were

XSL•FO
RenderX

found in the VR group compared to the control group. Improvement in the RCFT copy task was associated with a frontal-occipital functional connectivity increase revealed by rsfMRI in the VR group compared to the control group.

Conclusions: Fully immersive VR cognitive training had positive effects on the visuospatial function, apathy, affect, quality of life, and increased frontal-occipital functional connectivity in older people in a predementia state. Future trials using VR cognitive training with larger sample sizes and more sophisticated designs over a longer duration may reveal greater improvements in cognition, psychiatric symptoms, and brain functional connectivity.

Trial Registration: Clinical Research Information Service KCT0005243; https://tinyurl.com/2a4kfasa

(J Med Internet Res 2021;23(5):e24526) doi:10.2196/24526

KEYWORDS

virtual reality; cognitive training; visuospatial function; fMRI; visual network; mild cognitive impairment

Introduction

Dementia is a major neurodegenerative disorder, affecting approximately 10% of older people [1]. Cognitive, psychological, and behavioral deterioration are typical manifestations of dementia, ultimately resulting in functional impairments and disability [2]. The individual and societal burden of dementia is accelerating rapidly compared to other diseases [1,3]; however, due to the unclear mechanisms and multifactorial pathology underlying the development and progression of dementia, only symptomatic treatments are currently available [4].

To date, many researchers have suggested that prevention is crucial, and have identified risk and protective factors associated with dementia, as well as preventive strategies [5]. According to a recent large study, one-third of Alzheimer disease (AD) cases are attributable to potentially modifiable risk factors such as educational attainment, vascular factors, and depression [6]. Additionally, lifelong exposure to cognitively and mentally engaging activities has been shown to protect against cognitive decline [7], and performance of cognitively stimulating activities in advanced age was associated with better cognitive function [8]. Accordingly, recent cognitive training studies have shown that repeated practice of exercises to restore brain and cognitive reserves resulted in small to moderate positive improvements in cognition in patients with mild to moderate dementia [9].

Advances in computer sciences and information and communication technology (ICT) have resulted in increased availability and accessibility of computerized cognitive training. Although conclusive results have yet to be found, preliminary studies have reported improvements in trained and nontrained cognition, and enhanced brain activity in related regions after computerized cognitive training in individuals with mild cognitive impairment (MCI) [10-12]. Working memory training was effective in improving verbal memory and hippocampal activation in patients with MCI [11], and exposure to a driving video game resulted in increased ability to control the vehicle that was related to midline frontal theta power in older people [12]. Moreover, emerging ICT applications using virtual reality (VR) have resulted in evolutions in health care, including cognitive and behavioral therapy [13]. VR can offer interventions in flexible and real world–like environments, facilitating visuospatial function through learning and transference outcomes [14], highlighting a role for cognitive training in a virtual environment in basic research and clinical practice. Owing to the lack of knowledge and dearth of experiments on VR-based cognitive training, especially the fully immersive type [15,16], further studies are needed to ascertain its potential therapeutic efficacy.

Recently, the number of neuroimaging studies attempting to reveal the underlying neural mechanisms associated with cognitive decline has increased [17,18]. Functional connectivity studies using resting-state functional magnetic resonance imaging (rsfMRI) have identified networks temporally coinciding with spatially distant neurophysiological events that are intrinsically coherent during a resting state such as the default-mode network [19]. We considered that functional connectivity studies using rsfMRI may be able to reveal the neural mechanism, especially in the visual network, responsible for the observed cognitive improvements following VR cognitive training, as such training is based on the cognitive reserve hypothesis associated with functional neural networks [20].

To test this hypothesis, we performed a preliminary randomized controlled trial to determine the efficacy and mechanisms of VR cognitive training in a predementia state. We aimed to ascertain the effects of VR multidomain cognitive training on visuospatial function, comprehensive neuropsychological function, and psychiatric symptoms in predementia. Moreover, we examined the hypothesis that cognitive improvement could be related to increased functional connectivity in the visual network of the brain.

Methods

Participants

Participants over 60 years old in a predementia state (ranging from subjective cognitive decline to MCI) were prospectively recruited between May and December 2019 from the memory clinic of Gachon University Gil Medical Center, Republic of Korea. Among 58 individuals who were assessed for eligibility using structured clinical interviews and brain MRI, four participants were excluded due to cerebral infarction on MRI (n=2), severe white matter hyperintensity on MRI (n=1), and history of a recent dental implant surgery (n=1). Nine participants voluntarily withdrew from the study due to an acute medical condition (n=2), hospitalization of a family member (n=1), scheduling conflict (n=1), and unknown personal reasons

XSL•FO
RenderX

(n=5). Finally, a total of 45 participants were randomly assigned to either the VR group or the control group.

All participants had subjective cognitive complaints, including memory decline, but did not meet the criteria for diagnosis of a major neurocognitive disorder based on the Diagnostic and Statistical Manual of Mental Disorders (5th edition) [2]. Participants were classified as having subjective cognitive decline according to the corresponding research criteria and five cognitive domain scores such as attention, language, visuospatial, memory, and frontal executive functions above –1.5 SD [21,22]. Participants were classified as having MCI according to the Petersen criteria [23]. Screening evaluation of the participants was performed by a board-certified psychiatrist (JK) and a clinical neuropsychologist (SL).

The exclusion criteria for the participants were as follows: (i) Korean version of Mini-Mental State Examination (MMSE) score <20; (ii) impaired activities of daily living; (iii) comorbidity of severe medical or surgical conditions; (iv) major psychiatric disorders; (v) history of any kind of dementia; (vi) history of neurodegenerative disorders, including Creutzfeldt-Jakob disease, Pick disease, Huntington disease, Parkinson disease, inflammation associated with HIV, and syphilis; (vii) structural abnormalities on MRI such as intracranial hemorrhage, cerebral, cerebellar, or brainstem infarction, hydrocephalus, traumatic brain injury, severe white matter hyperintensity, tumors, multiple sclerosis, or vasculitis; and (viii) inability to use the VR system.

Information on study objectives, group allocation, cognitive intervention, brief study protocol, risks and benefits, and confidentiality was given to all participants before enrollment. All participants provided offline written informed consent, and the Institutional Review Board of Gachon University Gil Medical Center approved this study (GCIRB2018-396).

Study Design

This was an open-label, randomized controlled trial (KCT0005243) that aimed to investigate the efficacy of a fully immersive VR cognitive training program on visuospatial function in older people with risk for dementia (Multimedia Appendix 1). Participants were randomly assigned to either the VR or the control group. The unblinded randomization was performed by drawing lots with the participants present. Participants in both groups were evaluated for visuospatial function, comprehensive neuropsychological function, and psychiatric symptoms, and underwent rsfMRI before and after the 1-month cognitive training. The participants in the VR group underwent VR cognitive training twice a week for a total of eight sessions in addition to their usual therapy such as pharmacotherapy for the prevention of dementia (eg, choline alfoscerate and cholinesterase inhibitor); the participants in the control group did not undergo any additional intervention except for their usual therapy such as pharmacotherapy.

VR Cognitive Training

The multidomain VR cognitive training program was developed between November 2018 and April 2019 by the authors who are board-certified geriatric neuropsychiatrists and clinical neuropsychologists with expertise. The VR cognitive training program consisted of multiple games involving multidomain cognitive tasks to assess: (i) attention (to find differences), (ii) executive function and memory (to select items needed to perform certain tasks), (iii) working memory and ability to perform mathematical calculations (to prepare an exact amount of money), (iv) visuospatial orientation (to find a path using a memorized map), (v) visuospatial function (to spatially place furniture exactly based on a memorized drawing), (vi) verbal memory (to remember certain words), (vii) visual memory (to remember specific flags and symbols), and (viii) processing speed and working memory (to catch animals in a certain order). All virtual environments were fully immersive 3D settings allowing for feelings of increased presence and visuospatial stimulation; training was accompanied by game elements to increase the interest and motivation of the participants. Representative images of the VR training program are presented in Multimedia Appendix 2.

Each session lasted approximately 20-30 minutes. The VR training took place using a head-mounted Oculus Rift CV1 display, with Oculus Touch controllers held in both of the participant's hands. Each training session was performed with the participant in a seated position, and the difficulty level increased throughout the study period from easy to difficult (levels 1-4), with two sessions at each difficulty level. All procedures were performed in the memory clinic of Gachon University Gil Medical Center and were guided by a certified clinical neuropsychologist (SL) in addition to automatic verbal and visual messages from the program. There were no revisions, updates, or breaches of the program during the study period. This program was used exclusively in this study and is not available for commercial use.

Procedures and Outcome Measures

All participants underwent face-to-face comprehensive neuropsychological tests and evaluations using psychiatric scales, as well as rsfMRI at baseline and after the VR cognitive training period. Baseline evaluations of diagnostic criteria included global and functional scales such as the Korean version of the MMSE, Clinical Dementia Rating (CDR), CDR Sum of Boxes (CDR-SOB), global deterioration scale, and instrumental activities of daily living scales.

The primary outcome was the effect of the VR cognitive training on visuospatial function measured by the Rey-Osterrieth Complex Figure Test (RCFT) copy task, which has been validated in the Korean population [24,25]. Basic components, including a large rectangle (1 point), diagonal cross (1 point), horizontal midline of a large rectangle (1 point), and vertical midline of a large rectangle (1 point), were also evaluated because they are considered important in qualitative aspects [26,27]. The neuropsychologist (GP) who scored the RCFT copy task was blinded to the randomization.

The secondary outcomes concerned the effect on comprehensive neuropsychological function; psychiatric symptoms such as affect, apathy, quality of life (QoL), and depression; and functional connectivity in the visual network of the brain.

The neuropsychological tests consisted of the MMSE and subtests from the comprehensive neuropsychological test battery

XSL•FO
RenderX

[25]. Attention was assessed by the digit span forward and backward test and Trail Making Test (TMT) part A [25]. The Korean version of the Boston Naming Test (K-BNT) was used to assess language ability [25,28]. Memory was assessed by measuring performance on three tasks of the Seoul Verbal Learning Test (SVLT): immediate recall, delayed recall after 20 minutes, and recognition [25]. Frontal executive function was assessed by phonemic word fluency testing, the TMT-B, and the Stroop Color Test [25]. All neuropsychological test results were adjusted for age and years of education, and are presented as standardized *z*-scores.

Noncognitive psychiatric symptoms that typically start to decline in the early dementia stage were also assessed [29]. Depressive symptoms were evaluated by the validated 30-item Geriatric Depression Scale (GDS), including questions pertaining to mood, anxiety, energy, satisfaction, hopefulness, inattention, and sleep quality [30,31]. The GDS comprises a series of binary yes/no questions (scored as 1 or 0, respectively), with higher scores indicating severe depression. Apathy was evaluated by the validated 18-item Apathy Evaluation Scale (AES), including items pertaining to emotional affect, behavior, and cognitive apathy [32,33]. Items of the AES are rated on a 4-point Likert scale, with a low score indicating severe apathy. Affect was evaluated by the Positive and Negative Affect Schedule (PANAS), which consists of 10 items to assess positive affect (PANAS-P) measures such as alertness and enthusiasm and 10 items to assess negative affect (PANAS-N) such as lethargy and feelings of sadness [34,35]. Each of the PANAS items is rated from 1 (not at all) to 5 (very much), with higher scores indicating higher affect. Participants' QoL was evaluated by the QoL-AD scale, which has been validated for use in people with dementia, including 13 subjective rating items to assess physical health, living situation, relationships with friends, and the ability to engage in leisure activities [36,37]. Items of the QoL-AD are assessed on a 4-point Likert scale, with higher scores indicating better QoL.

The Simulator Sickness Questionnaire (SSQ) was administered after each session to evaluate tolerability of the VR cognitive training program [38]. Simulator sickness refers to side effects from virtual environment usage, and is also called cybersickness [39,40] and VR sickness [41]. The SSQ consists of 16 items yielding three subscales (nausea, oculomotor, and disorientation) and a total severity score, with high scores indicating increased symptoms. The levels of interest and satisfaction were also assessed on a Likert scale ranging from 0 to 100 after the period of VR cognitive training in a face-to-face manner.

MRI Acquisition

A 3-Tesla whole-body Siemens scanner (TrioTim syngo) was used for functional image acquisition with an interleaved T2*-weighted echo-planar imaging gradient echo sequence (repetition time/echo time=2500/25 milliseconds, flip angle=90°, slice thickness=3.5 mm, in-plane resolution=3.5×3.5 mm, matrix size=64×64) with a 12-channel birdcage head coil. For each participant, 160 functional volumes were acquired at the pretraining and posttraining time points. After rsfMRI, an anatomical image was acquired using a high T1-weighted 3D-gradient echo pulse sequence with magnetization-prepared rapid gradient echo (repetition time/echo time/inversion time=1900/3.3/900 milliseconds, flip angle=9°, slice thickness=1.0 mm, in-plane resolution=0.5×0.5 mm, matrix size=416×512). T1-weighted images were acquired only at the pretraining time point.

Functional Connectivity Analyses With rsfMRI

Preprocessing of the rsfMRI data was performed using Statistical Parametric Mapping software version 12 (Wellcome Trust Centre for Neuroimaging). First, a slice-timing correction was applied and the center of each image was relocated near the anterior commissure. Second, rsfMRI and T1-weighted images were imported into CONN FC toolbox v19c [42] for further preprocessing. To correct for between-scan rigid body motion, the functional images were realigned to the first image in the time series. The functional images were coregistered with anatomical images and spatially normalized to the Montreal Neurological Institute space using a transformation matrix derived from the T1-weighted anatomical image segmentation. The functional images were then resliced to 2×2×2 mm and spatially smoothed using an 8-mm full width at half maximum Gaussian kernel.

All preprocessed rsfMRI images were bandpass-filtered (0.008-0.09 Hz), and physiological and other spurious noise sources in the blood oxygenation level–dependent signal were removed using the anatomical component-based noise correction strategy implemented in CONN [43]. Outliers were calculated using the Artifact Detection Tools toolbox [44], and six motion correction parameters obtained from realignment were also modeled as nuisance covariates. The seed-to-voxel analyses were performed in the visual network with four cortical seed regions (right visual lateral, left visual lateral, visual medial, and visual occipital cortices) with predefined regions of interest based on the Harvard-Oxford atlas (fMRIB Software Library) [45]. Seed-based analyses were adjusted for age, years of education, sex, CDR-SOB, depressive symptoms, and pharmacotherapy. The mean time series for each seed region was calculated and then correlated with the time courses of all other voxels in the brain for each participant.

Sample Calculation and Statistical Analyses

Sample calculation was based on a recent meta-analysis on the effectiveness of VR for people with MCI or dementia that produced small-to-medium effect sizes using a random-effects model (effect size=0.29) from a total of 11 studies [15]. Assuming an attrition rate of 20%, a total sample size of 32 patients (16 per treatment group) would provide 0.8 power at a two-sided α error of .05. Power analysis was performed with G*Power software version 3.1.9.2.

Comparisons of demographic and clinical variables between the two groups were performed using independent *t* tests, the Mann-Whitney *U* test, or the χ^2 test. The paired *t* test was used in within-group comparisons of pretraining and posttraining measures. Repeated-measures analyses of variance was used to find the group interaction of the VR cognitive training on neuropsychological function and psychiatric symptom scales after adjusting for age, years of education, sex, CDR-SOB, depressive symptoms, and pharmacotherapy. Age and years of

XSL•FO
RenderX

education were not adjusted in analyses with comprehensive neuropsychological test results that are presented as age- and years of education–adjusted *z*-scores. All statistical analyses were performed with SPSS software version 23 (SPSS Inc), with a significance level assessed at $P<.05$ (two-tailed).

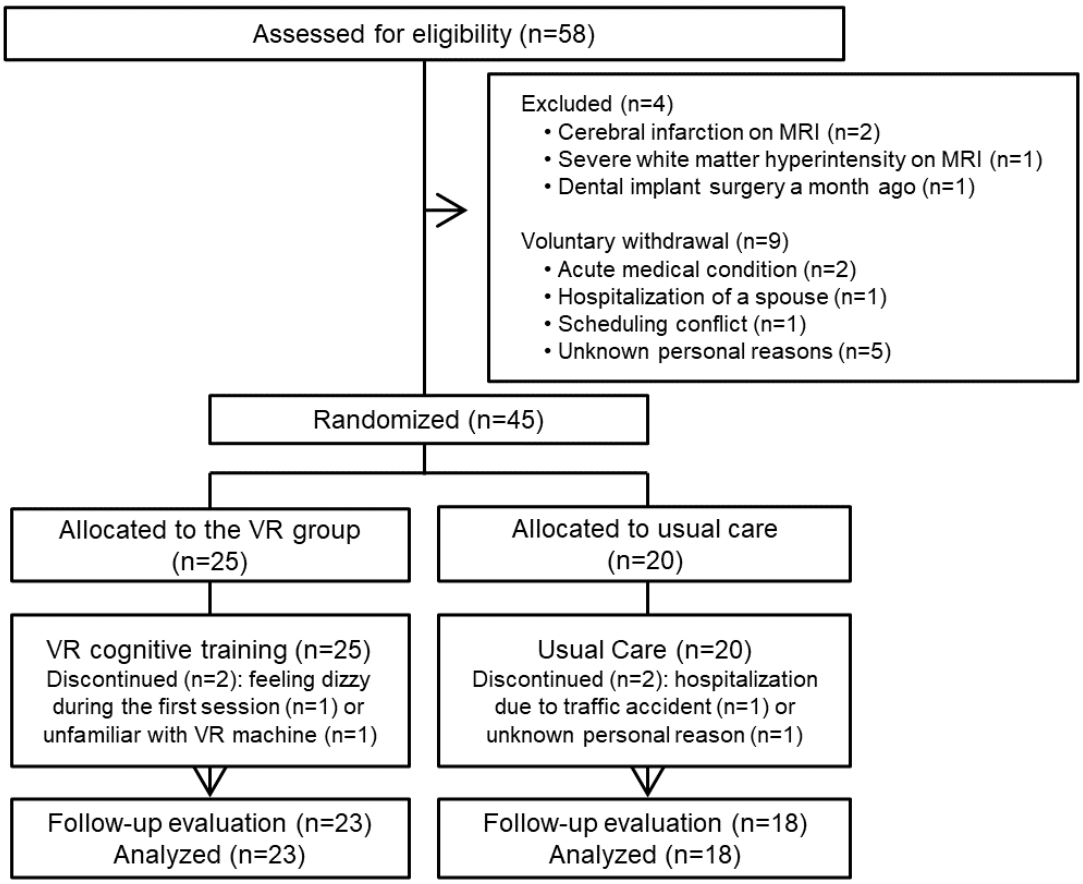
For rsfMRI data, Pearson correlation coefficients were converted to normally distributed scores using the Fisher *r*-to-*z* transformation. Group-level comparisons between the VR and control groups were performed using a general linear model in which improved cognitive task score was used as an explanatory variable and the posttraining minus pretraining *z*-transformation value was used as a dependent variable after adjusting for age, sex, years of education, CDR-SOB, depressive symptoms, and pharmacotherapy. The statistical thresholds for significance were set at voxel-wise uncorrected $P<.001$ and cluster-wise corrected $P<.05$ to correct for false-positive rates.

Results

Participants

Of the 45 participants who were randomly allocated to the VR (n=25) or the control (n=20) group, 41 participants completed the study. After allocation, two participants of the VR group dropped out of the study due to dizziness (n=1) and unfamiliarity with the VR machine during the first session (n=1). Two participants of the control group dropped out because of hospitalization due to a traffic accident (n=1) and unknown personal reasons (n=1). Ultimately, 41 participants were included in the analyses. The trial flow chart is presented in Figure 1.

Figure 1. Trial flow chart. VR: virtual reality; MRI: magnetic resonance imaging.



Demographic and Clinical Characteristics

Table 1 presents the detailed demographic and clinical characteristics of the study participants. Among the 41 participants, 23 (56%) and 18 (44%) were assigned to the VR and control groups, respectively. Participant age was around 75 years, and they were predominantly women. No group differences were found in the baseline diagnostic evaluation.

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of all study participants.

Characteristic	Total (N=41)	VR[a] group (n=23)	Control group (n=18)	χ^2 or U[b]	*P* value
Age (years), mean (SD)	74.51 (5.81)	75.48 (4.67)	73.28 (6.96)	–1.13	.26
Sex (female), n (%)	29 (70.7)	17 (73.9)	12 (66.7)	0.26	.61
Education (years), mean (SD)	8.07 (4.39)	7.70 (4.10)	8.56 (4.83)	–0.01	.99
MMSE[c], mean (SD)	26.24 (2.85)	26.22 (2.91)	26.28 (2.87)	0.09	.93
CDR[d], mean (SD)	0.41 (0.22)	0.41 (0.19)	0.42 (0.26)	–0.02	.99
CDR-SOB[e], mean (SD)	0.92 (1.00)	0.98 (0.85)	0.83 (1.19)	–1.24	.21
Global Deterioration Scale, mean (SD)	2.20 (0.78)	2.26 (0.75)	2.11 (0.83)	–0.78	.44
IADL[f], mean (SD)	0.13 (0.24)	0.14 (0.21)	0.11 (0.28)	–1.09	.27

[a]VR: virtual reality.

[b]Mann-Whitney *U* tests were used for all group comparisons except for sex, which was compared using the Pearson χ^2 test.

[c]MMSE: Mini-Mental State Examination.

[d]CDR: Clinical Dementia Rating.

[e]CDR-SOB: CDR-Sum of Boxes.

[f]IADL: instrumental activities of daily living.

Effect of VR Cognitive Training on Visuospatial Function

Table 2 shows the comparisons of the pretraining and posttraining visuospatial function within groups, as well as group interactions in the effects of VR cognitive training. VR training resulted in significant improvement in the RCFT copy task compared to the control group. Basic components of the RCFT copy task also improved in the VR group compared to the control group.

Table 2. Group comparisons of visuospatial function pre and post virtual reality (VR) cognitive training.

Function score	Pretraining	Posttraining	Within groups pretraining vs posttraining		Between groups interaction[a]		
			t (df)	*P* value	$F_{1,39}$	*P* value	η^2
RCFT[b] copy (*z*-score[c]), mean (SD)					14.69	.001	0.30
VR (n=23)	–0.31 (1.09)	0.22 (0.78)	–3.50 (22)	.002			
Control (n=18)	–0.07 (1.14)	–0.47 (1.22)	2.15 (17)	.046			
RCFT copy basic components[d], mean (SD)					9.27	.005	0.22
VR	1.99 (0.59)	2.14 (0.42)	–2.82 (22)	.01			
Control	2.15 (0.54)	2.07 (0.59)	1.53 (17)	.14			

[a]Repeated measures analysis of variance after adjusting for age (for basic components only), years of education (for basic components only), sex, Clinical Dementia Rating-Sum of Boxes, depressive symptoms, and pharmacotherapy.

[b]RCFT: Rey-Osterrieth Complex Figure Test; basic components consist of a large rectangle, diagonal cross, horizontal midline of a large rectangle, and vertical midline of a large rectangle.

[c]Adjusted for age and years of education.

[d]Raw scores.

Effect of VR Cognitive Training on Comprehensive Neuropsychological Function

Table 3 shows the comparisons of pretraining and posttraining comprehensive neuropsychological function within groups, as well as group interactions in the effects of VR cognitive training. K-BNT, SVLT delayed recall, and Controlled Oral Word Association Test phonemic fluency showed improvement in the VR group, but the group interaction was not significant.

XSL•FO
RenderX

Table 3. Group comparisons of comprehensive neuropsychological tests pre and post virtual reality (VR) cognitive training.

Test[a]	Pretraining, mean (SD)	Posttraining, mean (SD)	Within groups pretraining vs posttraining		Between groups interaction[b]		
			t (df)	*P* value	$F_{1,39}$	*P* value	η^2
Global cognition							
MMSE[c]					0.75	.39	0.02
VR (n=23)	26.22 (2.91)	25.87 (3.36)	0.97 (22)	.34			
Control (n=18)	26.28 (2.87)	26.67 (3.09)	–0.89 (17)	.39			
Attention							
Digit span, forward					0.00	.96	0.00
VR (n=23)	–0.11 (1.21)	–0.24 (0.87)	0.57 (22)	.57			
Control (n=18)	–0.08 (1.08)	0.15 (1.03)	–1.42 (17)	.18			
Digit span, backward					0.04	.84	0.00
VR (n=23)	–0.09 (0.99)	–0.15 (0.92)	0.23 (22)	.82			
Control (n=18)	–0.23 (1.26)	–0.25 (0.82)	0.08 (17)	.94			
TMT[d] -A					2.32	.14	0.06
VR (n=23)	0.13 (0.58)	0.12 (0.64)	0.10 (22)	.93			
Control (n=18)	–0.87 (4.19)	–0.38 (3.53)	–1.00 (17)	.33			
Language and related functions							
K-BNT[e]					3.55	.07	0.09
VR (n=23)	–0.23 (1.08)	0.19 (1.02)	–4.08 (22)	<.001			
Control (n=18)	–0.15 (1.00)	–0.01 (1.37)	–0.72 (17)	.48			
Verbal memory							
SVLT[f], immediate recall					1.83	.19	0.05
VR (n=23)	0.23 (0.10)	0.67 (1.24)	–3.10 (22)	.005			
Control (n=18)	0.30 (0.83)	0.52 (0.89)	–2.29 (17)	.04			
SVLT, delayed recall					3.03	.09	0.08
VR (n=23)	–0.10 (1.40)	0.66 (1.37)	–4.59 (22)	<.001			
Control (n=18)	0.12 (0.97)	0.58 (0.94)	–3.21 (17)	.005			
SVLT, recognition					0.37	.55	0.01
VR (n=23)	0.29 (1.39)	0.48 (1.30)	–0.93 (22)	.36			
Control (n=18)	0.37 (1.01)	0.29 (1.07)	0.41 (17)	.69			
Frontal executive function							
COWAT[g], semantic fluency					0.04	.85	0.00
VR (n=23)	–0.25 (0.99)	–0.44 (1.17)	1.01 (22)	.32			
Control (n=18)	–0.41 (1.00)	–0.58 (0.88)	1.09 (17)	.29			
COWAT, phonemic fluency					3.08	.09	0.08
VR (n=23)	–0.35 (0.88)	–0.41 (0.78)	0.39 (22)	.70			
Control (n=18)	–0.09 (0.82)	0.27 (1.01)	-1.89 (17)	.08			
Stroop test, color/word reading					0.05	.82	0.00
VR (n=23)	–0.01 (1.12)	0.32 (1.04)	–1.99 (22)	.06			
Control (n=18)	–0.01 (0.85)	0.16 (1.21)	–0.64 (17)	.53			
TMT-B, mean (SD)					0.13	.73	0.00

XSL•FO
RenderX

Test[a]	Pretraining, mean (SD)	Posttraining, mean (SD)	Within groups pretraining vs posttraining		Between groups interaction[b]		
			t (df)	*P* value	$F_{1,39}$	*P* value	η^2
VR (n=23)	–1.43 (2.04)	–0.64 (1.74)	–2.30 (22)	.03			
Control (n=18)	–0.55 (1.52)	–0.55 (1.62)	0.01 (17)	.996			

[a]All data except for MMSE are presented as age and years of education–adjusted *z*-scores.

[b]Repeated-measures analysis of variance after adjusting for age (for MMSE only), years of education (for MMSE only), sex, Clinical Dementia Rating-Sum of Boxes, depressive symptoms, and pharmacotherapy.

[c]MMSE: Mini-Mental State Examination.

[d]TMT-B: Trail Making Test.

[e]K-BNT: Korean version of the Boston Naming Test.

[f]SVLT: Seoul Verbal Learning Test.

[g]COWAT: Controlled Oral Word Association Test.

Effect of VR Cognitive Training on Psychiatric Symptoms

Table 4 shows the comparisons between the pretraining and posttraining measures based on psychiatric symptoms within groups, as well as group differences in the effects of VR cognitive training. Group differences were found in the AES, PANAS-P, PANAS-N, and QoL-AD measures, showing improvements in apathy, positive and negative affect, and QoL in the VR group.

Table 4. Group comparisons of psychiatric symptoms pre and post virtual reality (VR) cognitive training.

Group	Pretraining, mean (SD)	Posttraining, mean (SD)	Within groups pretraining vs posttraining		Between groups interaction[a]		
			t (df)	*P* value	$F_{1,39}$	*P* value	η^2
GDS[b]					0.88	.36	0.03
VR (n=23)	15.00 (6.08)	13.26 (6.49)	2.46 (22)	.02			
Control (n=18)	12.17 (6.85)	11.72 (7.18)	0.47 (17)	.65			
AES[c]					7.02	.01	0.17
VR (n=23)	47.43 (10.20)	54.35 (9.41)	–3.04 (22)	.006			
Control (n=18)	52.83 (9.38)	51.22 (8.72)	0.98 (17)	.34			
PANAS-P[d]					14.40	.001	0.30
VR (n=23)	17.00 (6.28)	21.43 (7.27)	–2.71 (22)	.01			
Control (n=18)	21.83 (7.48)	16.50 (6.51)	4.63 (17)	<.001			
PANAS-N[e]					4.23	.047	0.11
VR (n=23)	18.22 (7.09)	16.30 (6.35)	0.97 (22)	.34			
Control (n=18)	18.89 (5.31)	20.44 (8.42)	–1.16 (17)	.26			
QoL-AD[f]					4.49	.04	0.12
VR (n=18)	31.04 (4.69)	32.26 (4.96)	–1.23 (22)	.23			
Control (n=23)	34.94 (9.43)	32.72 (6.54)	1.21 (17)	.25			

[a]Repeated-measures analysis of variance after adjusting for age, years of education, sex, Clinical Dementia Rating-Sum of Boxes, and pharmacotherapy.

[b]GDS: Geriatric Depression Scale.

[c]AES: Apathy Evaluation Scale.

[d]PANAS-P: Positive and Negative Affect Schedule-positive affect.

[e]PANAS-N: Positive and Negative Affect Schedule-negative affect.

[f]QoL-AD: Quality of Life-Alzheimer Disease.

XSL•FO
RenderX

Simulator Sickness, Interest, and Satisfaction Associated with the VR Training Program

Table 5 shows the simulator sickness measured by the SSQ after each training session, reported on a Likert scale ranging from 0 to 100, in the VR group participants after the training period. The mean SSQ total score was 12.86 (SD 11.82), and the summary subscale mean score for nausea, oculomotor, and disorientation was 7.02 (SD 6.40), 11.15 (10.56), and 17.16 (16.91), respectively.

Interest and satisfaction had mean scores of 79.78 (SD 14.18) and 78.04 (SD 12.50) on a Likert scale ranging from 0 to 100, respectively.

Table 5. Mean (SD) simulator sickness questionnaire scores associated with the virtual reality cognitive training (n=23).

Session	Nausea	Oculomotor	Disorientation	Total score
1	9.95 (14.80)	14.83 (14.18)	22.39 (22.89)	17.24 (17.53)
2	6.22 (10.61)	11.53 (16.30)	15.13 (18.25)	12.20 (15.59)
3	9.13 (12.69)	11.21 (16.62)	16.95 (25.51)	13.66 (18.86)
4	6.64 (9.73)	9.89 (14.89)	10.29 (19.78)	10.24 (14.97)
5	6.64 (10.54)	7.91 (13.02)	13.92 (23.37)	10.24 (14.63)
6	2.45 (7.17)	9.56 (15.70)	20.58 (26.83)	11.22 (16.18)
7	5.39 (6.94)	9.89 (12.61)	19.37 (27.12)	12.20 (14.93)
8	9.97 (11.20)	14.82 (17.27)	18.98 (27.35)	16.32 (19.47)

Increased Functional Connectivity in rsfMRI

We investigated brain functional connectivity in the visual network associated with the improvement in the RCFT copy task. The areas with significantly increased connectivity in the seed-to-voxel visual networks are presented in Table 6: (a) from the right visual lateral cortices to the left paracingulate gyrus, right paracingulate gyrus, left frontal pole, left superior frontal gyrus, anterior cingulate gyrus, and white matter; and (b) from the visual medial cortices to the right insular cortex, right frontal pole, right frontal operculum cortex, right caudate, left caudate, right putamen, left insular cortex, and white matter.

Figure 2 depicts the increased regional functional connectivity in the brain cortices and the white matter that are related to improvements in the RCFT copy task in the VR group compared to the control group.

XSL•FO
RenderX

Table 6. Functional visual network connectivity related to improved Rey-Osterrieth Complex Figure Test copy task scores after virtual reality cognitive training.

Seed and connected regions (voxels)	Clusters		
	Voxel (2×2×2)	MNI[a] coordinates (x, y, z)[b]	FDR[c]–corrected *P* value[d]
Visual lateral, R[e]	291	–06, +40, +42	.003
Paracingulate gyrus, L[f]	118		
Paracingulate gyrus, R	68		
Frontal pole, L	41		
Superior frontal gyrus, L	29		
Anterior cingulate gyrus	7		
Frontal pole, R	1		
White matter	27		
Visual medial	719 and 401	+16, +20, +16 and –22, +22, +16	<.001 and <.001
Insular cortex, R	71		
Frontal pole, R	48		
Frontal operculum cortex, R	25		
Caudate, R	24		
Caudate, L	3		
Putamen, R	2		
White matter	546		
Insular cortex, L	2		
White matter	399		
Visual lateral, L	N/A[g]	N/A	N/A
Visual occipital	N/A	N/A	N/A

[a]MNI: Montreal Neurological Institute.

[b]Coordinates indicate the representative coverage region with maximum power among all connected regions.

[c]False-discovery Rate.

[d]Group-level analyses between the VR and control groups were performed using a general linear model with Rey-Osterrieth Complex Figure Test copy task improvement as an explanatory variable and the post-pre training *z* transformation value as a dependent variable after adjusting for age, years of education, sex, Clinical Dementia Rating-Sum of Boxes, depressive symptoms, and pharmacotherapy.

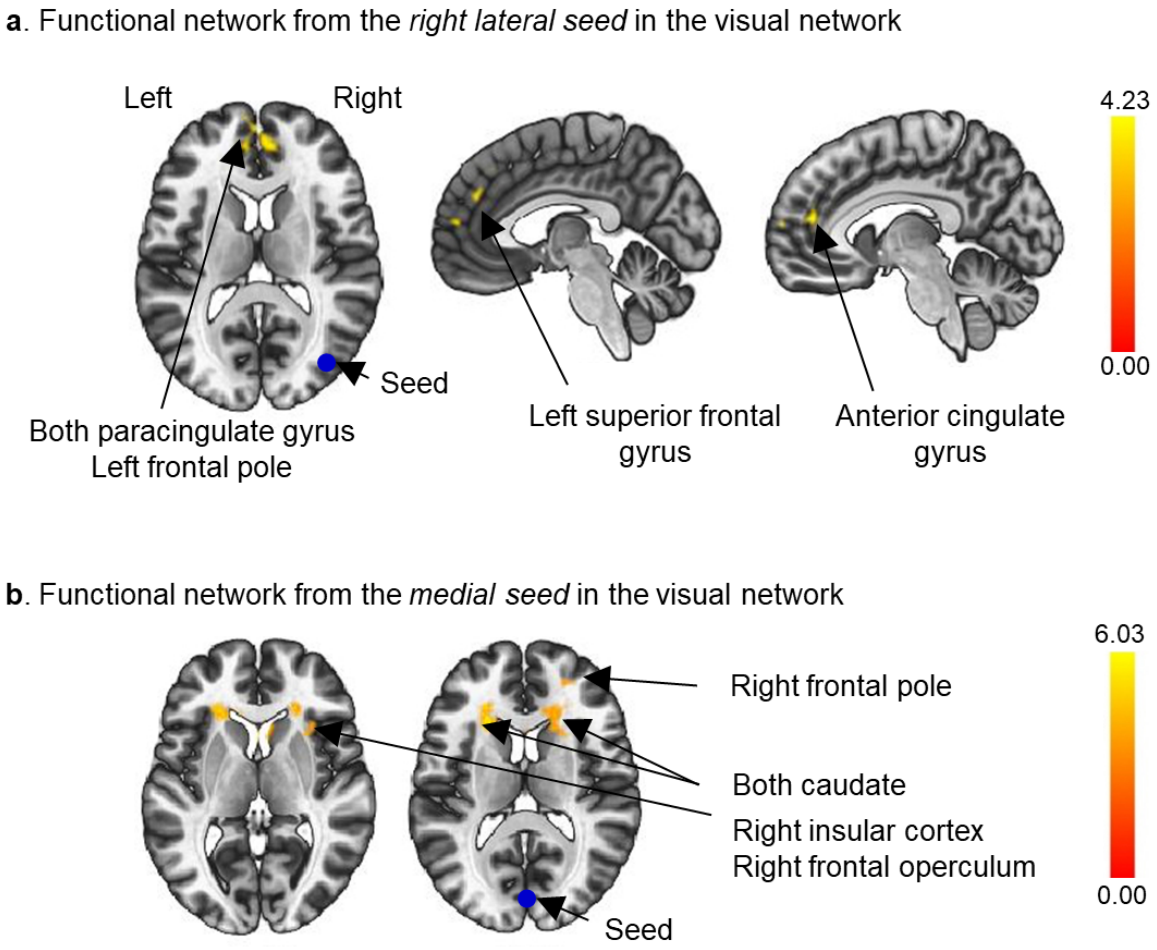
[e]R: right side.

[f]L: left side.

[g]N/A: not applicable.

XSL•FO
RenderX

Figure 2. Seed-to-voxel analyses based on the right lateral region (a) and the medial region (b) of the visual network (blue circles). Increased frontal-occipital functional connectivity related to the Rey-Osterrieth Complex Figure Test copy task improvement after virtual reality cognitive training. False discovery rate–corrected *P*<.05 for cluster threshold; uncorrected *P*<.001 for voxel threshold.



Discussion

Principal Findings

This study found that 1-month multidomain cognitive training using fully immersive VR was effective in improving visuospatial function and frontal-occipital functional connectivity, as well as apathy, affect, and QoL in older people in a predementia cognitive state.

The first major finding of this study is that VR cognitive training resulted in improvements in the RCFT copy task. Despite the inconsistent results reported in the literature, training-related changes in cognition have been repeatedly found in older people with cognitive disorders [9,10]. Neuropsychological test score improvements after traditional pen-and-paper or computerized cognitive training have been found in measures of global composite cognition [46-48], verbal memory [11,46,49,50], verbal letter fluency [46,47], verbal fluency [51,52], and visuospatial function in the clock-drawing test [46,53]. It has also been reported that VR cognitive training was effective in improving frontal executive function in individuals with MCI [54], as well as attention and visual memory in older people [55,56]. In line with these previous studies, our results also showed that multidomain cognitive training in a virtual environment was effective in visuospatial function measured by the RCFT copy task in total and basic components comprising the gestalt of the feature showing the ability to approach [26,27]. Although not significant, improvements in naming ability, verbal memory delayed recall, and frontal executive function were also found in the training group compared to the control group. It is possible that the relatively short 1-month training period might have resulted in the lack of group difference, as a learning effect may have impacted the posttraining neuropsychological test results in the control group. However, the improvement in visuospatial function in the VR group, even after the short period of cognitive training, might be attributed to the ecological nature of the fully immersive VR environment. In the enriched auditorily and visually stimulating environment, processing of visual orientation, visuospatial construction, and visual selective attention likely occurred [57,58]. In recent studies with VR evaluation, investigators have been able to effectively differentiate between the navigational [59] and visuospatial deficits observed in patients with MCI and healthy older people [60,61]. In studies with VR interventions, VR cognitive training was found to be effective [55,62] or ineffective [56,63] in improving visuospatial function

XSL•FO
RenderX

in older people or those in an early dementia stage. We believe that the cognitive training performed in the maximally immersive environment with the head-mounted display, headphones, and hand movement trackers in our study might have increased visuospatial functioning in those at a predementia stage [64]. The immersion methods utilized in previous studies investigating VR cognitive training in older people have employed desktop-based systems [55,56], screen and sensors [62], screen and glasses [65,66], and head-mounted display and fixed joystick setups [63]. Although heterogeneity in study populations and methodological differences among prior studies have resulted in inconsistent findings, this study provides further evidence to support the benefits of VR cognitive training in eliciting improvements in visuospatial processing through the repeated presentation of real-world, dynamic, multisensory, and interactive environments.

Another novel finding was the increased functional connectivity observed in the frontal-occipital cortical network after VR cognitive training, which was associated with improved performance in the RCFT copy task, consistent with the associations between cognitive improvements and neuronal plasticity that have been observed previously [67]. In patients with MCI, significant associations have been observed between verbal memory improvement and left hippocampal activation in task-related fMRI after 8-week training to improve auditory processing speed and accuracy [11]. Other studies have shown that 6 weeks of episodic memory training in patients with MCI resulted in the manifestation of new associations between improved delayed word recall test performance and brain activation in the right inferior parietal lobule in fMRI during memory encoding [68]. In healthy older people, 8 weeks of exposure to a cognitive control training program led to an increased frontoparietal network related to cognitive control ability [69]; another study found that verbal recall was associated with an increased left hippocampal volume in healthy older people after 8 weekly verbal recall memory training sessions [70]. Thus, in this study, repetitive cognitive training in a novel fully immersive environment might have increased the frontal-occipital activation in accordance with improved visuospatial function. We also observed increased functional connectivity in white matter areas, which are known to exhibit a lower hemodynamic response than the grey matter. Although fMRI studies have focused on grey matter until recently, the increased functional connectivity in the white matter close to the grey matter supports the growing neural evidence of fMRI white matter changes induced by VR cognitive training [71,72].

This evident link between visuospatial construction and frontal-occipital functional connectivity might be explained by the acquired cognitive system engagement induced by the RCFT copy task, which requires the participant to copy a complex geometric figure [73]. Visuoconstructive ability is based on the Van Sommers model of drawing [74]; according to this cognitive model, the RCFT copy task consists of (i) visual recognition of a 2D Rey-Osterrieth complex figure; (ii) visual representation of the figure in long-term or temporary memory; (iii) graphical output processes such as those related to depiction decisions (eg, context, orientation, viewpoint, details, and boundary) or reproduction strategies (eg, copying orders, dimensions, shapes, diagonals, crosses, line sets); (iv) graphical planning (eg, routine or contingent planning); and (v) articulation and economic constraints during motor output. Through these steps, multiple brain regions have been found to be associated with performance in the RCFT copy task, including the temporal, parietal, occipital, and frontal cortices in both hemispheres or in the right hemisphere alone [75-77]. Although we observed increased activity only in the primary visual cortices (visual medial) and the right associative visual cortices (right visual lateral) connecting to the areas in the middle frontal cortices, these regions are known to be involved in the visual recognition and graphic output planning processes required to complete the RCFT copy task [74], and are associated with visuo-motor transformation and multistep object use in the task [77]. A recent study reported that lesions in the right superior parietal lobe and the middle occipital gyrus were associated with poor RCFT copy task performance [78], which is in accordance with our results. Furthermore, there have been reports on improvements in nontrained cognitive functions, also known as transfer effects, in memory training in older people with MCI [79,80]. Previous studies have shown that repeated memory-focused training might have enhanced the processing speed of memory retrieval and the efficiency of working memory, assuming that frontal executive function was the main recipient of the transfer effects [79,80]. Although recent studies have applied cognitive training with novel computerized tools and involvement of multiple cognitive domains, existing programs have only applied cognitive training in a 2D environment with an emphasis on language abilities [9,53,79,81]. Since frontal executive function plays a major role in all cognitive domains and higher-order cognitive controls [82], the improved performance on the RCFT copy task may be supported by increased functional connectivity in the frontal-occipital network.

The psychiatric benefit of VR cognitive training in individuals in a predementia state should be considered. In this study, participants in the VR group showed improved apathy, affect, and QoL scores after training compared with those in the control group. A recent review reported that computerized cognitive training resulted in long-term improvements in psychological outcome measures [16]. Although methodologies vary across studies, 3D VR cognitive training was effective in improving depressive symptoms in patients with MCI compared with an active control group receiving music therapy [63]. Moreover, a few feasibility studies have reported improved alertness, pleasure, apathy, and security following one-time exposure to a less immersive VR environment [65,66]. We postulate that apathy, affect, and QoL might be improved by the VR cognitive training, as these are some of the early symptoms of dementia [83]. Immersive virtual environments might facilitate the limited functioning of patients with cognitive disorders that affect communication, interaction, motivation, engagement, and positive attitudes toward others [84]. Thus, the importance of virtual environments should be considered in cognitive training because the feeling of presence itself in a 3D space can enhance volitional motivation, allowing one to constantly process external stimuli and cognitively adjust to changing environments [85].

XSL•FO
RenderX

Simulator sickness reported after every session was minimal in the VR cognitive training group. In this study, the SSQ total score (mean 12.86, SD 11.82) indicated minimal symptoms (score 5-10) according to the suggested categorization established in flight simulators [86]. Although the SSQ was originally developed and validated in military personnel using flight simulators, it is the most commonly used measure of sickness in a virtual environment [40]. A recent meta-analysis of the SSQ in virtual environments reported total and subscale scores and dropout rates according to VR conditions (total 28.00, SD 1.71; nausea 16.72, SD 0.77; oculomotor 17.09, SD 0.55; disorientation 23.50, SD 1.17; dropout rate 15.6%) [87]. Compared to the results in a VR environment, our results on the SSQ scores and dropout rates (8%) showed better tolerability, with moderate interest and satisfaction. The SSQ scores have been reported to be higher in VR than in a flight simulator environment [86] and with gaming content than without it [87]. Despite the characteristics of VR in this study, such as fully immerse, game and training components, and old users at an early stage of cognitive decline, these results may imply that fully immersive VR can be a safe and interesting method for cognitive training.

Limitations

Our study had several strengths and limitations. This is one of the largest VR cognitive training studies to use a fully immersive 3D VR program. Compared to 2D or semi-immersive VR programs, our results highlight the positive effects of employing fully immersive 3D VR in cognitive training, as we found neural evidence supporting the improvement in visuospatial function. However, there are several limitations and lessons learned in this study. First, the small sample size and short training period were the main limitations. Although sample sizes in studies investigating the effects of cognitive training are increasing [88], most VR trials still rely on small sample sizes and are performed over a short duration, especially those using fully virtual environments [15]. Short clinical trial periods in previous studies investigating the effect of computerized cognitive training programs have also been a limiting factor in the field as a whole [88]. Thus, future studies should aim to increase the sample sizes and extend the duration of training to better evaluate the effect of VR cognitive training. Second, we considered that the per-protocol analysis could bias the results of this randomized controlled trial, although the number of participants who dropped out of the study was the same in both groups. Third, the lack of an active control group in this study is another limitation. Some previous trials have included active control groups receiving psychoeducation, cognitive therapy, face-to-face music therapy, or pen-and-paper cognitive training for comparisons with the VR training group [15,63]. In the future, various active control groups should be considered to confirm the effectiveness of VR cognitive training. Fourth, the lack of examination for AD biomarkers such as cerebrospinal fluid analysis or brain imaging for amyloid detection can be a limitation because it is unclear whether the participants in our study will develop AD, which is the most prevalent cause of dementia. Future studies involving AD biomarkers could clearly explain the pure effect of cognitive training in individuals in a preclinical or prodromal dementia state. Lastly, heterogeneity among patients, practitioners, program content, and accessibility to the VR system can limit the generalizability of the results to other populations.

Conclusions

We found that fully immersive VR cognitive training improved cognition and psychiatric symptoms in a predementia state. Visuospatial function improved in such individuals relative to controls, and this finding was supported by increased frontal-occipital functional connectivity assessed by rsfMRI. These findings suggest that VR training can enhance visuospatial ability by exposing patients to an enriched virtual environment, leading to improved apathy, affect, and QoL. Our results support the neurotherapeutic use of VR cognitive training as an effective nonpharmacological intervention for those who are at risk for dementia; however, more rigorous trials should be performed to confirm the effects and identify the associated neural mechanisms.

Acknowledgments

We would like to thank C2MONSTER for help in designing and manufacturing the multidomain VR cognitive training program. This research was funded by the support program for Development of Dementia Care Service using Advanced ICT Technology 2018 funded by Korea Radio Promotion Association (RAPA) and the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center support program (IITP-2021-2017-0-01630) supervised by the Institute for Information & Communications Technology Promotion. The funding sources had no role in the study design; collection, analysis, and interpretation of data; writing of the manuscript; or decision to submit the article for publication.

Authors' Contributions

JK, NK, and SL conceived and designed the study, acquired and analyzed the data, interpreted the study findings, and drafted the manuscript. SW, GP, and JP analyzed the data. BY, JL, JY, and SR designed the study, interpreted the study findings, supervised and directed the conduct of the study, and critically reviewed the manuscript. SC conceived and designed the study, acquired and analyzed the data, interpreted the study findings, supervised and directed the conduct of the study, and critically reviewed the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
CONSORT-EHEALTH (V 1.6.1) - Submission/Publication Form.
[PDF File (Adobe PDF File), 446 KB - jmir_v23i5e24526_app1.pdf]

Multimedia Appendix 2
Representative images of the virtual reality training program.
[DOCX File , 1988 KB - jmir_v23i5e24526_app2.docx]

References

1. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. Alzheimers Dement 2016 Apr;12(4):459-509. [doi: 10.1016/j.jalz.2016.03.001] [Medline: 27570871]
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, DC: American Psychiatric Pub; 2013.
3. 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. JAMA 2018 Apr 10;319(14):1444-1472 [FREE Full text] [doi: 10.1001/jama.2018.0158] [Medline: 29634829]
4. Cao J, Hou J, Ping J, Cai D. Advances in developing novel therapeutic strategies for Alzheimer's disease. Mol Neurodegener 2018 Dec 12;13(1):64 [FREE Full text] [doi: 10.1186/s13024-018-0299-8] [Medline: 30541602]
5. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]
6. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 2014 Aug;13(8):788-794. [doi: 10.1016/S1474-4422(14)70136-X] [Medline: 25030513]
7. Marioni RE, Proust-Lima C, Amieva H, Brayne C, Matthews FE, Dartigues J, et al. Cognitive lifestyle jointly predicts longitudinal cognitive decline and mortality risk. Eur J Epidemiol 2014 Mar 28;29(3):211-219 [FREE Full text] [doi: 10.1007/s10654-014-9881-8] [Medline: 24577561]
8. Wilson RS, Segawa E, Boyle PA, Bennett DA. Influence of late-life cognitive activity on cognitive health. Neurology 2012 Apr 10;78(15):1123-1129 [FREE Full text] [doi: 10.1212/WNL.0b013e31824f8c03] [Medline: 22491864]
9. Bahar-Fuchs A, Martyr A, Goh A, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. Cochrane Database Syst Rev 2019 Mar 25;3:CD013069 [FREE Full text] [doi: 10.1002/14651858.CD013069.pub2] [Medline: 30909318]
10. Gates N, Vernooij R, Di Nisio M, Karim S, March E, Martínez G, et al. Computerised cognitive training for preventing dementia in people with mild cognitive impairment. Cochrane Database Syst Rev 2019 Mar 13;3:CD012279 [FREE Full text] [doi: 10.1002/14651858.CD012279.pub2] [Medline: 30864747]
11. Rosen AC, Sugiura L, Kramer JH, Whitfield-Gabrieli S, Gabrieli JD. Cognitive training changes hippocampal function in mild cognitive impairment: a pilot study. J Alzheimers Dis 2011;26(Suppl 3):349-357 [FREE Full text] [doi: 10.3233/JAD-2011-0009] [Medline: 21971474]
12. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 05;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
13. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 03;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
14. Biocca F. The cyborg's dilemma: progressive embodiment in virtual environments 1. J Comput Mediat Commun 1997;3(2):JCMC324. [doi: 10.1111/j.1083-6101.1997.tb00070.x]
15. Kim O, Pang Y, Kim J. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. BMC Psychiatry 2019 Jul 12;19(1):219 [FREE Full text] [doi: 10.1186/s12888-019-2180-x] [Medline: 31299921]
16. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
17. Sperling R. Potential of functional MRI as a biomarker in early Alzheimer's disease. Neurobiol Aging 2011 Dec;32(Suppl 1):S37-S43 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2011.09.009] [Medline: 22078171]
18. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 2011 Dec;32(12):2322-2327 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2010.05.023] [Medline: 20594615]
19. Buckner R, Andrews-Hanna J, Schacter D. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci 2008 Mar;1124:1-38. [doi: 10.1196/annals.1440.011] [Medline: 18400922]

XSL•FO
RenderX

20. Stern Y. Cognitive reserve. Neuropsychologia 2009 Aug;47(10):2015-2028 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2009.03.004] [Medline: 19467352]
21. Molinuevo JL, Rabin LA, Amariglio R, Buckley R, Dubois B, Ellis KA, Subjective Cognitive Decline Initiative (SCD-I) Working Group. Implementation of subjective cognitive decline criteria in research studies. Alzheimers Dement 2017 Mar 05;13(3):296-311. [doi: 10.1016/j.jalz.2016.09.012] [Medline: 27825022]
22. Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, Subjective Cognitive Decline Initiative (SCD-I) Working Group. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014 Nov 02;10(6):844-852 [FREE Full text] [doi: 10.1016/j.jalz.2014.01.001] [Medline: 24798886]
23. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
24. Meyers JE, Meyers KR. Rey complex figure test under four different administration procedures. The Clinical Neuropsychologist 1995 Feb;9(1):63-67. [doi: 10.1080/13854049508402059]
25. Kang Y, Jahng S, Na D. Seoul Neuropsychological Screening Battery, 2nd Edition (SNSB-II). Seoul: Human Brain Research & Consulting Co; 2012.
26. Bylsma F, Bobholz J, Schretlen D, Correa D. A brief, reliable approach to coding how subjects copy the Rey-Osterrieth Complex Figure. J Int Neuropsychol Soc 1995;1:125.
27. Sargénius HL, Bylsma FW, Lydersen S, Hestad K. Visual-constructional ability in individuals with severe obesity: Rey Complex Figure Test accuracy and the Q-Score. Front Psychol 2017 Sep 20;8:1629. [doi: 10.3389/fpsyg.2017.01629] [Medline: 28979230]
28. Kim H, Na DL. Normative data on the Korean version of the Boston Naming Test. J Clin Exp Neuropsychol 1999 Feb 09;21(1):127-133. [doi: 10.1076/jcen.21.1.127.942] [Medline: 10421007]
29. Lyketsos CG, Carrillo MC, Ryan JM, Khachaturian AS, Trzepacz P, Amatniek J, et al. Neuropsychiatric symptoms in Alzheimer's disease. Alzheimers Dement 2011 Sep 01;7(5):532-539 [FREE Full text] [doi: 10.1016/j.jalz.2011.05.2410] [Medline: 21889116]
30. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. J Psychosom Res 2004 Sep;57(3):297-305. [doi: 10.1016/j.jpsychores.2004.01.004] [Medline: 15507257]
31. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. J Psych Res 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
32. Marin RS, Biedrzycki RC, Firinciogullari S. Reliability and validity of the apathy evaluation scale. Psych Res 1991 Aug;38(2):143-162. [doi: 10.1016/0165-1781(91)90040-v]
33. Lee YM, Park IH, Koo MS, Ko SY, Kang HM, Song JE. The reliability and validity of the Korean version of Apathy Evaluation Scale and its application in patients with schizophrenia. Korean J Schizophr Res 2013;16(2):80. [doi: 10.16946/kjsr.2013.16.2.80]
34. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Person Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
35. Lim Y, Yu B, Kim D, Kim J. The positive and negative affect schedule: psychometric properties of the korean version. Psychiatry Investig 2010 Sep;7(3):163-169 [FREE Full text] [doi: 10.4306/pi.2010.7.3.163] [Medline: 20927304]
36. Thorgrimsen L, Selwood A, Spector A, Royan L, de Madariaga Lopez M, Woods RT, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. Alzheimer Dis Assoc Disord 2003;17(4):201-208. [doi: 10.1097/00002093-200310000-00002] [Medline: 14657783]
37. Shin H. A preliminary study on the Korean version of quality of life-Alzheimer's disease (QOL-AD) scale in community-dwelling elderly with dementia. J Prev Med Public Health 2006 May;39(3):243-248 [FREE Full text] [Medline: 16764499]
38. 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]
39. LaViola JJ. A discussion of cybersickness in virtual environments. SIGCHI Bull 2000 Jan;32(1):47-56. [doi: 10.1145/333329.333344]
40. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virt Real 2016 Apr 26;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
41. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
42. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect 2012 Jun;2(3):125-141. [doi: 10.1089/brain.2012.0073] [Medline: 22642651]
43. Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 2007 Aug 01;37(1):90-101 [FREE Full text] [doi: 10.1016/j.neuroimage.2007.04.042] [Medline: 17560126]
44. Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage 2010 Aug 15;52(2):571-582 [FREE Full text] [doi: 10.1016/j.neuroimage.2010.04.246] [Medline: 20420926]

XSL•FO
RenderX

45. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004 Jan;23(Suppl 1):S208-S219. [doi: 10.1016/j.neuroimage.2004.07.051] [Medline: 15501092]
46. Mapelli D, Di Rosa E, Nocita R, Sava D. Cognitive stimulation in patients with dementia: randomized controlled trial. Dement Geriatr Cogn Dis Extra 2013;3(1):263-271 [FREE Full text] [doi: 10.1159/000353457] [Medline: 24052800]
47. Bergamaschi S, Arcara G, Calza A, Villani D, Orgeta V, Mondini S. One-year repeated cycles of cognitive training (CT) for Alzheimer's disease. Aging Clin Exp Res 2013 Aug;25(4):421-426. [doi: 10.1007/s40520-013-0065-2] [Medline: 23784727]
48. Jeong JH, Na HR, Choi SH, Kim J, Na DL, Seo SW, et al. Group- and home-based cognitive intervention for patients with mild cognitive impairment: a randomized controlled trial. Psychother Psychosom 2016 May 27;85(4):198-207 [FREE Full text] [doi: 10.1159/000442261] [Medline: 27230861]
49. Cavallo M, Angilletta C. Long-lasting neuropsychological effects of a computerized cognitive training in patients affected by early stage Alzheimer's disease: are they stable over time? J Appl Gerontol 2019 Jul;38(7):1035-1044. [doi: 10.1177/0733464817750276] [Medline: 29295652]
50. Jelcic N, Cagnin A, Meneghello F, Turolla A, Ermani M, Dam M. Effects of lexical-semantic treatment on memory in early Alzheimer disease: an observer-blinded randomized controlled trial. Neurorehabil Neural Repair 2012 Oct;26(8):949-956. [doi: 10.1177/1545968312440146] [Medline: 22460609]
51. Trebbastoni A, Imbriano L, Podda L, Rendace L, Sacchetti ML, Campanelli A, et al. Cognitive training in patients with Alzheimer's disease: findings of a 12-month randomized controlled trial. Curr Alzheimer Res 2018 Mar 14;15(5):452-461. [doi: 10.2174/1567205014666171113105044] [Medline: 29141548]
52. Cavallo M, Hunter EM, van der Hiele K, Angilletta C. Computerized structured cognitive training in patients affected by early-stage Alzheimer's disease is feasible and effective: a randomized controlled study. Arch Clin Neuropsychol 2016 Dec 01;31(8):868-876. [doi: 10.1093/arclin/acw072] [Medline: 27600448]
53. Nousia A, Siokas V, Aretouli E, Messinis L, Aloizou A, Martzoukou M, et al. Beneficial effect of multidomain cognitive training on the neuropsychological performance of patients with early-stage Alzheimer's disease. Neural Plast 2018 Jul 11;2018:2845176-2828459. [doi: 10.1155/2018/2845176] [Medline: 30123243]
54. Liao Y, Chen I, Lin Y, Chen Y, Hsu W. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019 Jul 16;11:162. [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
55. Gamito P, Oliveira J, Morais D, Coelho C, Santos N, Alves C, et al. Cognitive stimulation of elderly individuals with instrumental virtual reality-based activities of daily life: pre-post treatment study. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):69-75. [doi: 10.1089/cyber.2017.0679] [Medline: 30040477]
56. Gamito P, Oliveira J, Alves C, Santos N, Coelho C, Brito R. Virtual reality-based cognitive stimulation to improve cognitive functioning in community elderly: controlled study. Cyberpsychol Behav Soc Netw 2020 Mar 01;23(3):150-156. [doi: 10.1089/cyber.2019.0271] [Medline: 32031888]
57. Green CS, Bavelier D. Action video game modifies visual selective attention. Nature 2003 May 29;423(6939):534-537. [doi: 10.1038/nature01647] [Medline: 12774121]
58. West GL, Stevens SA, Pun C, Pratt J. Visuospatial experience modulates attentional capture: evidence from action video game players. J Vis 2008 Dec 22;8(16):13.1-9 [FREE Full text] [doi: 10.1167/8.16.13] [Medline: 19146279]
59. Cushman LA, Stein K, Duffy CJ. Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. Neurology 2008 Sep 16;71(12):888-895 [FREE Full text] [doi: 10.1212/01.wnl.0000326262.67613.fe] [Medline: 18794491]
60. Zygouris S, Giakoumis D, Votis K, Doumpoulakis S, Ntovas K, Segkouli S, et al. Can a virtual reality cognitive training application fulfill a dual role? using the Virtual Supermarket Cognitive Training Application as a screening tool for mild cognitive impairment. J Alzheimers Dis 2015 Feb 19;44(4):1333-1347. [doi: 10.3233/jad-141260]
61. Howett D, Castegnaro A, Krzywicka K, Hagman J, Marchment D, Henson R, et al. Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. Brain 2019 Jun 01;142(6):1751-1766. [doi: 10.1093/brain/awz116] [Medline: 31121601]
62. Tarnanas I, Schlee W, Tsolaki M, Müri R, Mosimann U, Nef T. Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study. JMIR Serious Games 2013 Aug 06;1(1):e1 [FREE Full text] [doi: 10.2196/games.2778] [Medline: 25658491]
63. Optale G, Urgesi C, Busato V, Marin S, Piron L, Priftis K, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010 May 24;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
64. Sayma M, Tuijt R, Cooper C, Walters K. Are we there yet? Immersive virtual reality to improve cognitive function in dementia and mild cognitive impairment. Gerontologist 2020 Sep 15;60(7):e502-e512. [doi: 10.1093/geront/gnz132] [Medline: 31600389]

XSL•FO
RenderX

65. Manera V, Chapoulie E, Bourgeois J, Guerchouche R, David R, Ondrej J, et al. A feasibility study with image-based rendered virtual reality in patients with mild cognitive impairment and dementia. PLoS One 2016 Mar 18;11(3):e0151487 [FREE Full text] [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
66. Moyle W, Jones C, Dwan T, Petrovich T. Effectiveness of a virtual reality forest on people with dementia: a mixed methods pilot study. Gerontologist 2018 May 08;58(3):478-487. [doi: 10.1093/geront/gnw270] [Medline: 28329867]
67. Ten Brinke LF, Davis JC, Barha CK, Liu-Ambrose T. Effects of computerized cognitive training on neuroimaging outcomes in older adults: a systematic review. BMC Geriatr 2017 Jul 10;17(1):139 [FREE Full text] [doi: 10.1186/s12877-017-0529-x] [Medline: 28693437]
68. Belleville S, Clément F, Mellah S, Gilbert B, Fontaine F, Gauthier S. Training-related brain plasticity in subjects at risk of developing Alzheimer's disease. Brain 2011 Jun;134(Pt 6):1623-1634. [doi: 10.1093/brain/awr037] [Medline: 21427462]
69. Kim H, Chey J, Lee S. Effects of multicomponent training of cognitive control on cognitive function and brain activation in older adults. Neurosci Res 2017 Nov;124:8-15 [FREE Full text] [doi: 10.1016/j.neures.2017.05.004] [Medline: 28577979]
70. Engvig A, Fjell AM, Westlye LT, Skaane NV, Dale AM, Holland D, et al. Effects of cognitive training on gray matter volumes in memory clinic patients with subjective memory impairment. J Alzheimers Dis 2014;41(3):779-791. [doi: 10.3233/JAD-131889] [Medline: 24685630]
71. Gawryluk JR, Mazerolle EL, D'Arcy RCN. Does functional MRI detect activation in white matter? A review of emerging evidence, issues, and future directions. Front Neurosci 2014;8:239. [doi: 10.3389/fnins.2014.00239] [Medline: 25152709]
72. Grajauskas LA, Frizzell T, Song X, D'Arcy RCN. White matter fMRI activation cannot be treated as a nuisance regressor: overcoming a historical blind spot. Front Neurosci 2019 Oct 4;13:1024. [doi: 10.3389/fnins.2019.01024] [Medline: 31636527]
73. Shin M, Park S, Park S, Seol S, Kwon JS. Clinical and empirical applications of the Rey-Osterrieth Complex Figure Test. Nat Protoc 2006 Jul 27;1(2):892-899. [doi: 10.1038/nprot.2006.115] [Medline: 17406322]
74. Sommers PV. A system for drawing and drawing-related neuropsychology. Cognitive Neuropsychol 1989 Mar;6(2):117-164. [doi: 10.1080/02643298908253416]
75. Melrose RJ, Harwood D, Khoo T, Mandelkern M, Sultzer DL. Association between cerebral metabolism and Rey-Osterrieth Complex Figure Test performance in Alzheimer's disease. J Clin Exp Neuropsychol 2013 Feb 07;35(3):246-258 [FREE Full text] [doi: 10.1080/13803395.2012.763113] [Medline: 23387510]
76. Kang SH, Park YH, Lee D, Kim JP, Chin J, Ahn Y, et al. The cortical neuroanatomy related to specific neuropsychological deficits in Alzheimer's continuum. Dement Neurocogn Disord 2019 Sep;18(3):77-95 [FREE Full text] [doi: 10.12779/dnd.2019.18.3.77] [Medline: 31681443]
77. Chen H, Pan X, Lau JKL, Bickerton W, Pradeep B, Taheri M, et al. Lesion-symptom mapping of a complex figure copy task: A large-scale PCA study of the BCoS trial. Neuroimage Clin 2016;11:622-634 [FREE Full text] [doi: 10.1016/j.nicl.2016.04.007] [Medline: 27182489]
78. Biesbroek JM, van Zandvoort MJE, Kuijf HJ, Weaver NA, Kappelle LJ, Vos PC, Utrecht VCI Study Group. The anatomy of visuospatial construction revealed by lesion-symptom mapping. Neuropsychologia 2014 Sep;62:68-76. [doi: 10.1016/j.neuropsychologia.2014.07.013] [Medline: 25062545]
79. Youn J, Park S, Lee J, Cho S, Kim J, Ryu S. Cognitive improvement in older adults with mild cognitive impairment: evidence from a multi-strategic metamemory training. J Clin Med 2020 Jan 28;9(2):362 [FREE Full text] [doi: 10.3390/jcm9020362] [Medline: 32013035]
80. Weng W, Liang J, Xue J, Zhu T, Jiang Y, Wang J, et al. The transfer effects of cognitive training on working memory among Chinese older adults with mild cognitive impairment: a randomized controlled trial. Front Aging Neurosci 2019 Aug 14;11:212. [doi: 10.3389/fnagi.2019.00212] [Medline: 31474850]
81. Han JW, Son KL, Byun HJ, Ko JW, Kim K, Hong JW, et al. Efficacy of the Ubiquitous Spaced Retrieval-based Memory Advancement and Rehabilitation Training (USMART) program among patients with mild cognitive impairment: a randomized controlled crossover trial. Alzheimers Res Ther 2017 Jun 06;9(1):39 [FREE Full text] [doi: 10.1186/s13195-017-0264-8] [Medline: 28587629]
82. Badre D. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn Sci 2008 May;12(5):193-200. [doi: 10.1016/j.tics.2008.02.004] [Medline: 18403252]
83. Lyketsos CG, Lopez O, Jones B, Fitzpatrick AL, Breitner J, DeKosky S. Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. JAMA 2002 Sep 25;288(12):1475-1483. [doi: 10.1001/jama.288.12.1475] [Medline: 12243634]
84. Garcia L, Kartolo A, Méthot-Curtis E. A discussion of the use of virtual reality in dementia. In: Virtual reality in psychological, medical and pedagogical applications. Rijeka: Intech; 2012:123-136.
85. Riva G, Waterworth JA, Waterworth EL, Mantovani F. From intention to action: The role of presence. New Ideas Psychol 2011 Jan;29(1):24-37. [doi: 10.1016/j.newideapsych.2009.11.002]
86. Kennedy R, Drexler J, Compton D, Stanney K, Lanham D, Harm D. Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences. In: Hettinger LJ, Haas MW, editors. Virtual and adaptive environments: Applications, implications, and human performance issues. Mahwah, NJ: Lawrence Erlbaum Associates; 2003:247.

XSL•FO
RenderX

87. Saredakis D, Szpak A, Birckhead B, Keage H, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. Front Hum Neurosci 2020;14:96. [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
88. Gates NJ, Rutjes AW, Di Nisio M, Karim S, Chong L, March E, et al. Computerised cognitive training for 12 or more weeks for maintaining cognitive function in cognitively healthy people in late life. Cochrane Database Syst Rev 2020 Feb 27;2:CD012277 [FREE Full text] [doi: 10.1002/14651858.CD012277.pub3] [Medline: 32104914]

Abbreviations

AD: Alzheimer disease
AES: Apathy Evaluation Scale
CDR: Clinical Dementia Rating
CDR-SOB: Clinical Dementia Rating-Sum of Boxes
fMRI: functional magnetic resonance imaging
GDS: Geriatric Depression Scale
ICT: information and communication technology
K-BNT: Korean version of the Boston Naming Test
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MRI: magnetic resonance imaging
PANAS: Positive and Negative Affect Schedule
PANAS-N: Positive and Negative Affect Schedule-negative
PANAS-P: Positive and Negative Affect Schedule-positive
QoL: quality of life
RCFT: Rey-Osterrieth Complex Figure Test
rsfMRI: resting-state functional magnetic resonance imaging
SSQ: Simulator Sickness Questionnaire
SVLT: Seoul Verbal Learning Test
TMT: Trail Making Test
VR: virtual reality

Edited by G Eysenbach; submitted 23.09.20; peer-reviewed by N Martin; comments to author 15.10.20; revised version received 26.10.20; accepted 11.04.21; published 06.05.21.

Please cite as:
Kang JM, Kim N, Lee SY, Woo SK, Park G, Yeon BK, Park JW, Youn JH, Ryu SH, Lee JY, Cho SJ
Effect of Cognitive Training in Fully Immersive Virtual Reality on Visuospatial Function and Frontal-Occipital Functional Connectivity in Predementia: Randomized Controlled Trial
J Med Internet Res 2021;23(5):e24526
URL: https://www.jmir.org/2021/5/e24526
doi:10.2196/24526
PMID:33955835

©Jae Myeong Kang, Nambeom Kim, Sook Young Lee, Soo Kyun Woo, Geumjin Park, Byeong Kil Yeon, Jung Woon Park, Jung-Hae Youn, Seung-Ho Ryu, Jun-Young Lee, Seong-Jin Cho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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 the Feasibility of an Open-Source Virtual Reality Mirror Visual Feedback Module for Complex Regional Pain Syndrome: Pilot Usability Study

Andrea Stevenson Won[1], PhD; Ariana C Barreau[2], MSc; Mark Gaertner[3], MD; Tristan Stone[4], BA; Joshua Zhu[5], BA; Cheng Yao Wang[5], MSc; Sean Mackey[6], MD, PhD

[1]Department of Communication, Cornell University, Ithaca, NY, United States
[2]Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Urbana, IL, United States
[3]Department of Neurology, University of California, Los Angeles, Los Angeles, CA, United States
[4]Department of Computer Science, Cornell University, Ithaca, NY, United States
[5]Department of Information Science, Cornell University, Ithaca, NY, United States
[6]Stanford Neuroscience and Pain Lab, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Andrea Stevenson Won, PhD
Department of Communication
Cornell University
471 Mann Library Building
Ithaca, NY, 14853
United States
Phone: 1 919 906 0287
Fax: 1 607 254 2790
Email: asw248@cornell.edu

Abstract

Background: Complex regional pain syndrome (CRPS) is a rare and severe chronic pain condition, with effective treatment options not established for many patients. The underlying pathophysiology remains unclear, but there is a growing appreciation for the role of central mechanisms which have formed the basis for brain-based therapies such as transcranial magnetic stimulation and mirror visual feedback (MVF). MVF has been deployed in the treatment of CRPS using both conventional mirrors and virtual reality (VR).

Objective: The aim of this study was to further investigate the use of VR in the treatment of patients with unilateral upper limb CRPS. VR has the potential advantage of more flexible and more motivating tasks, as well as the option of tracking patient improvement through the use of movement data.

Methods: We describe the development, acceptability, feasibility, and usability of an open-source VR program MVF module designed to be used with consumer VR systems for the treatment of CRPS. The development team was an interdisciplinary group of physical therapists, pain researchers, and VR researchers. Patients recruited from a pain clinic completed 3-5 visits each to trial the system and assessed their experiences in pre- and post-treatment questionnaires.

Results: All 9 (100%) participants were able to use the system for 3, 4, or 5 trials each. None of the participants quit any trial due to cybersickness. All 9 (100%) participants reported interest in using the module in the future. Participants' reported average pain scores in the affected limb were not significantly different from baseline during treatment or after treatment (P=.16). We did not find a statistically significant effect on participants' self-reported average pain scores.

Conclusions: We propose that this module could be a useful starting point for modification and testing for other researchers. We share modifications to make this module usable with standalone headsets and finger tracking. Next steps include adapting this module for at-home use, or for use with participants with lower limb pain.

(J Med Internet Res 2021;23(5):e16536) doi:10.2196/16536

KEYWORDS

virtual reality; pain; complex regional pain syndrome; CRPS; open source; mirror visual feedback

XSL•FO
RenderX

Introduction

Complex regional pain syndrome (CRPS) is a rare and severe chronic pain condition, defined as "continuing pain...disproportionate to any inciting event," accompanied by an array of varying signs and symptoms (sensory, vasomotor, sudomotor/edema, or motor/trophic features) which do not fit other diagnoses [1-3]. A Cochrane Review found low level of quality of evidence for CRPS treatments [4], with current treatment guidelines based largely on expert experience, case reports, open-label trials, and pilot studies [3]. Thus, many patients are without established, effective treatment options [5]. While the underlying pathophysiology remains unclear, there is a growing appreciation for the role of central mechanisms which have formed the basis for brain-based therapies such as transcranial magnetic stimulation [6] and mirror visual feedback (MVF). MVF [7] has been deployed in the treatment of CRPS using both conventional mirrors [8] and virtual reality (VR) [9-11]. VR has the advantage of more flexible and more motivating tasks [10-12], the ability to be used at home [13], and the option of tracking patient improvement through the use of movement data [11,12,14]. However, barriers of cost and difficulty of adoption have prevented this technology from being widely used.

Given the advent of inexpensive, portable consumer VR systems, systematically testing the potential of these systems for the treatment of CRPS and other disorders has become a possibility for many researchers who have not previously used VR in medicine. Virtual environments for therapeutic purposes can be created using game engines such as Unity 3D or Unreal. These virtual environments can be modified to allow researchers to investigate the effects of avatar realism, limb swapping, changing the appearance of the targets, changing the environment around the targets, etc. However, because many researchers working with clinical populations have limited experience building virtual environments, there is a significant barrier to entry.

To provide a tool for the research community as well as for our own research, we developed an open-source VR mirror module that could be used for our own experiments. We deployed this module in a pilot study for patients with chronic unilateral upper limb CRPS.

This pilot study was a feasibility study to test the usability of the VR system and its acceptance, ease of use, and patient willingness to engage. We evaluated these goals by examining participant retention, as well as qualitative comments on general usability and interest in engaging with the module at home. Participants' reported average pain scores in the affected limb were not significantly different from baseline after treatment (P=.16). Furthermore, all participants completed the required sessions for this study and qualitative measures indicated that participants were interested in continuing the therapy at home.

Methods

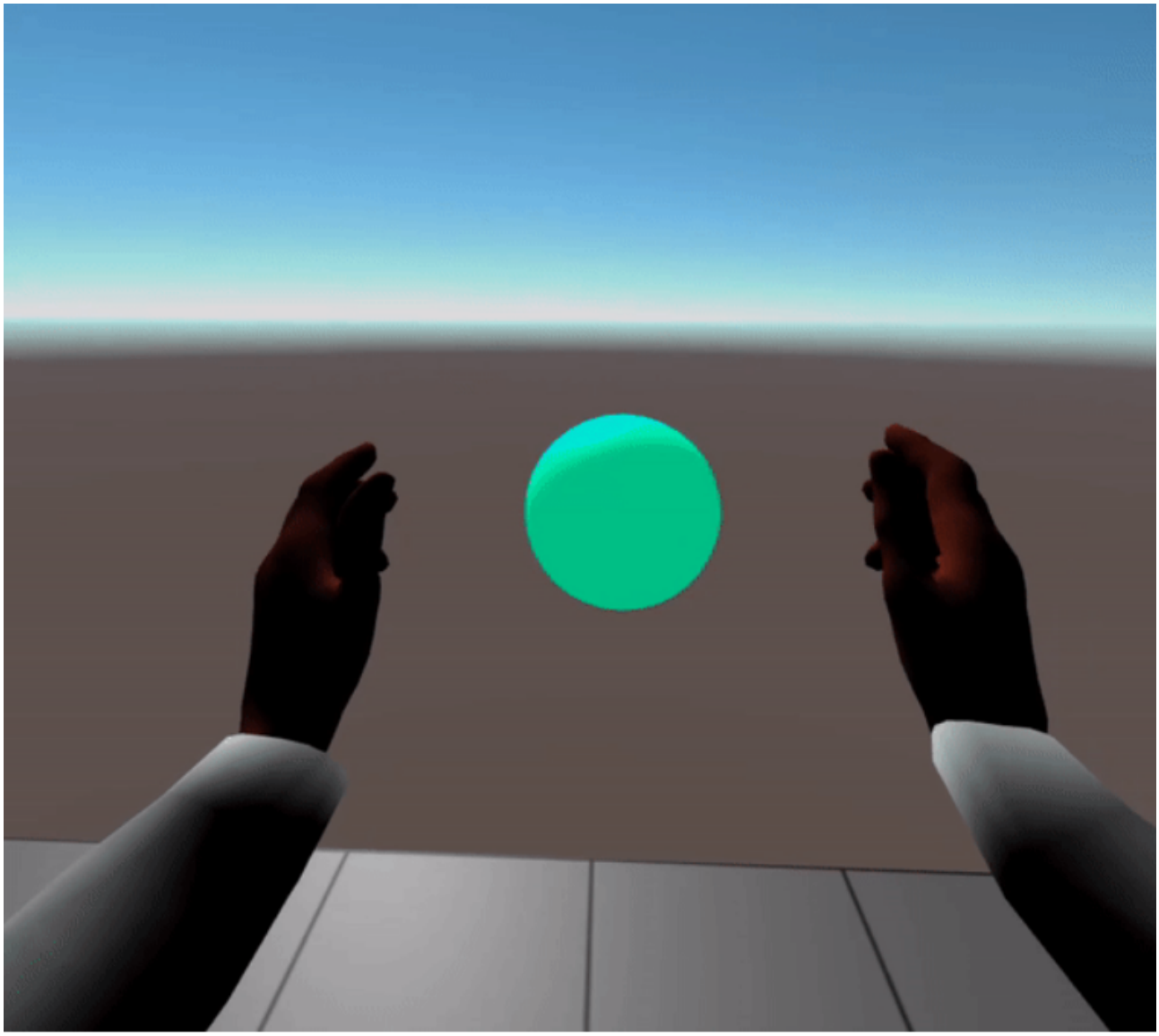
Virtual Reality Application

A team of pain researchers and VR researchers created a simple open-source VR application for patients with CRPS using the game engine Unity 3D. This application is compatible with a range of consumer VR systems that use head-mounted displays and hand controllers or hand tracking to allow users to interact with virtual content by controlling avatars. Figure 1 shows an example of a user's view of their avatar's hands which are controlled by their movements.

As in conventional MVF, movement from the patient's uninjured hand was transformed over the midline to symmetrically animate their avatar hand on the injured side. Thus, patients' uninjured hands could control the movements of both avatar hands, just as in conventional mirror therapy.

XSL•FO
RenderX

Figure 1. Image representing participant in VELOS MVF with a mono version of their first person perspective.



Participants

In the summer of 2017, 110 adult patients in the database of a pain clinic at a large United States university with a diagnosis of unilateral CRPS in an upper extremity were contacted to ask if they would be interested in participating in a pilot study to test the feasibility of a novel VR intervention. Eligible patients had average pain level reported on a Visual Analog Scale as greater than or equal to 4 out of 10, were on a stable treatment plan for the month preceding the study, and could commit to multiple visits. A total of 9 patients (6 female and 3 male) were recruited within the 2-month time span set for the study. Among the 9 patients, 7 (78%) identified as Caucasian, 1 (11%) as Asian, and 1 as “other” (11%). Their ages ranged from 19 to 60 years, with the median being 42 and mean 44. Their education level ranged from a high-school diploma to graduate degrees, and their income ranged from under US $10,000 to US $80,000 or more. All 9 patients were right handed, with 3 having CRPS on the left upper extremity and 6 with right upper extremity CRPS. All patients signed informed consent in the clinic, received US $20 per hour of visit time, and the experiment was approved by the Stanford Institutional Review Board. The consent form, along with all other study materials, is shown in Multimedia Appendix 1.

Table 1 shows participants’ symptom onset and duration.

XSL•FO
RenderX

Table 1. Participants' CRPS symptom onset and duration.

Patient ID	Symptom onset date	Duration (from the onset of CRPS[a] to the baseline assessment date)
2	August 16, 2008	8 years and 11 months
3	May 15, 2010	7 years and 2 months
4	December 12, 2012	4 years and 7 months
7	January 11, 2015	2 years and 6 months
8	June 1, 2013	4 years and 2 months
9	September 6, 2016	10 months and 29 days
10	August 7, 2016	1 year and 16 days
11	March 28, 2013	4 years and 4 months
12	February 2, 2015	4 years and 6 months

[a]CRPS: complex regional pain syndrome.

Materials

To make this environment as accessible to as many users as possible, it was built to accommodate both the hardware systems currently supported by the Steam VR plug-in in the Unity game engine. At the time of the experiment, these were the Oculus Rift/Touch system [14] and the HTC VIVE system [15], both of which cost less than US $400 and could be deployed using a "VR-ready" laptop as seen in Multimedia Appendix 2. The application can accommodate, but does not require, room-scale tracking. Since this initial study, the project has been revised to be used with standalone headsets that include hand and finger tracking. Original and revised projects and documentation can be found in [16].

Baseline Survey

All visits took place in July and August of 2017. After the screening interview, participants completed a baseline survey at the clinic through the institution's REDCap account. This survey consisted of a set of demographic and baseline measures, including the PROMIS measures of pain and daily function [17]. This survey was repeated after the last session, and 1 month after the study. All survey instruments are provided in Multimedia Appendix 1.

Procedure

Participants then returned to the clinic weekly for a minimum of 4 sessions of immersive VR, with an optional fifth visit based on participant interest and availability. After each session, participants completed a daily pain survey (modified from the Brief Pain Inventory [18]) and a presence questionnaire derived from Witmer and Singer [19].

Participants were seated in a swivel chair in front of the computer and a researcher (AB) helped them to don the headset and hold the hand controller in their unaffected hand. Avatars were scaled to the participants' seated height and the color of the arms and hands was selected at the beginning of each session. For participants suffering from CRPS of the right upper limb, both avatar hands were controlled using the left controller. For participants suffering from CRPS of the left upper limb, this was reversed and both avatar hands were controlled using the right controller. It was too painful for some participants to hold the tracker with their injured hand, so all participants held only 1 tracker on their uninjured side.

The researcher (AB) then began the session by pressing the spacebar key. Each time the key was pressed, one of a series of 18 targets of 3 different sizes (6 large, 6 medium, and 6 small) appeared in random order in the midline of the participant's field of view. Participants were asked to bring both the injured and uninjured hands together to contact the target. When an avatar hand contacted the target, a chime sounded, and the target object disappeared. Participants completed as many sets of target hitting as they wished (mean 7.5 [SD 2.25]). For the first 2 participants, the protocol differed slightly for their first 2 visits, in that they were only allowed to complete 2 sets.

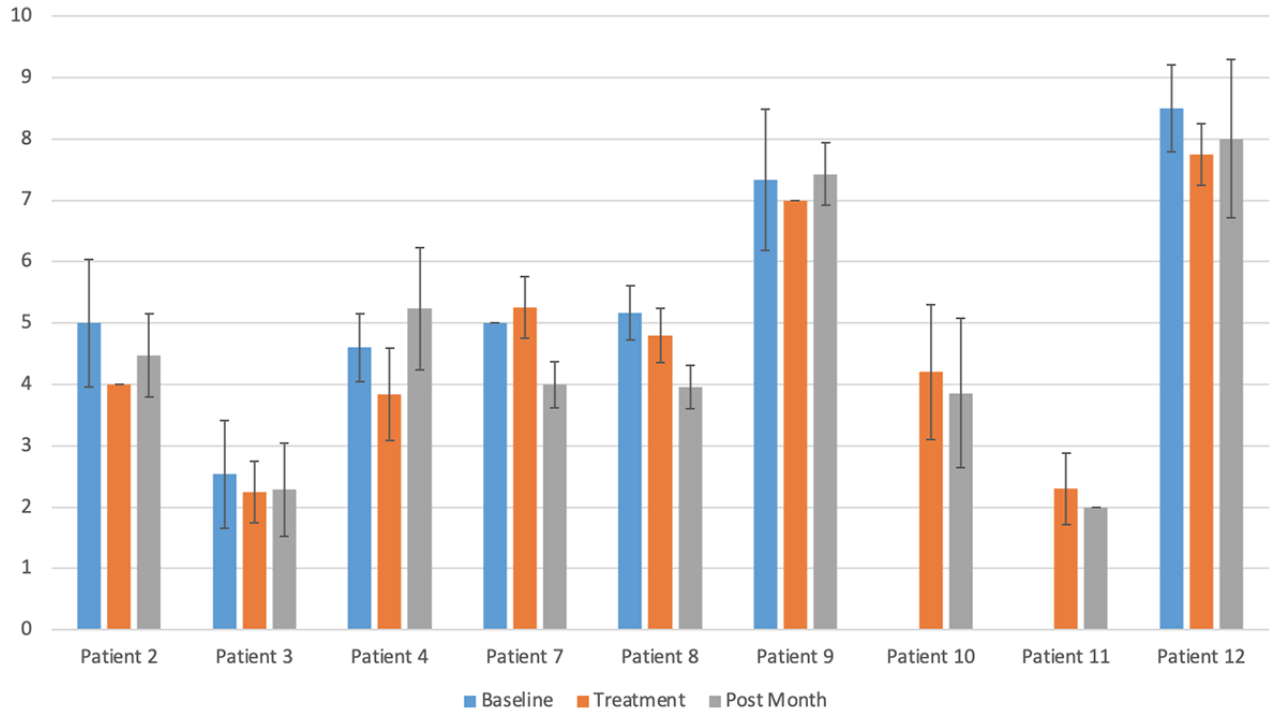
Measures Taken at Each Session

Pain, Physical Activity, Mood, and Quality of Sleep

Before completing the target-hitting task, participants reported their average level of pain throughout their body and in their affected CRPS limb, their level of physical activity, their mood, and the quality of their sleep over the past 24 hours, as well as their CRPS-related pain at that moment and any qualitative responses (Figure 2). These data are described in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 2. Patient ratings of average pain in affected CRPS limb. Ratings were on a 0-10 scale, with 10 being the most pain. Ratings were averaged before treatment, during treatment, and at one month post-treatment.



Movement Data

Participants' movements were recorded at a rate of 30 times per second. As each target was hit, the time and the position of the trackers at the time of contact were also recorded.

Cybersickness Measures

When considering the effectiveness of such a system in a clinical setting, it is important to assess its usability. Thus, we included a measure of cybersickness, also known as simulator sickness, which occurs when exposure to a virtual environment causes symptoms similar to motion sickness.

Results

All 9 participants completed 3, 4, or 5 sessions, and all were interested in continuing the therapy at home. Detailed results are presented in the following sections.

Baseline Survey

The results of the baseline surveys can be found in Multimedia Appendix 3.

Pain, Physical Activity, Mood, and Quality of Sleep

Several participants commented positively on the sessions, for example "VR helped hand pain!" and "I'm more likely to use my affected arm than I was before the study. Even my family noticed." However, using a linear mixed model with Bonferroni-corrected α of .01 for multiple comparisons, there were no statistically significant differences over time on average or highest pain of the affected limb or body, or on physical activity, mood, or quality of sleep. Table 2 shows these results.

Table 2. Nonsignificant effects of time on patient self-reported outcomes using linear mixed models, predicting each outcome measure averaged at each time point and including participant ID as a random effect. Alpha adjusted to .007 using Bonferroni correction for multiple comparisons. We note that this study was not powered to detect efficacy.

Outcome measure	*P* value
How would you rate your average level of pain in your affected CRPS limb over the past 24 hours?	.13
How would you rate your average level of pain in your throughout your body over the past 24 hours?	.76
How would you rate your highest level of pain in your affected CRPS limb over the past 24 hours?	.04
How would you rate your highest level of pain throughout your body over the past 24 hours?	.64
How would you rate your level of physical activity over the past 24 hours?	.75
Rate your overall mood for the past 24 hours.	.60
How would you rate the quality of your sleep over the past 24 hours?	.84

XSL•FO
RenderX

Movement Data

Most consumer VR systems contain trackers on the head-mounted display and 2 hand controllers, each of which captures both position (X,Y,Z) and orientation (pitch, yaw, roll) data at each frame. The default setting in our current module is to record every third frame, for a rate of approximately 30 frames per second. However, the frame rate can be modified to record at a higher or lower rate. We saved movement data for each tracker to a separate CSV file at the end of the experiment. We note that the data we recorded reflect the actual positions of the single tracker held by the participants, and the head-mounted display that they wore on their heads.

In this pilot study, participants' head and uninjured hand movements were tracked. Participants were not asked to hold or attach a hand sensor to their injured hand, on the grounds that this might exacerbate their injury. Thus, the data collected from these sensors cannot inform us about how participants may or may not have moved their injured arm during the experiment. Data collected from the other 2 sensors can be used to examine changes in movement over the course of the following sessions, and can also be compared with other, self-reported measures. A representative example of movement data can be found in Multimedia Appendix 4.

Cybersickness

The majority (7 participants) reported no cybersickness, with 2 participants rating their cybersickness as "slight." No participants quit any trial due to its effects.

Discussion

Principal Findings

In this pilot study, we tested the feasibility, ease of use, and patient willingness to engage with an open-source MVF VR module for CRPS. All 9 participants were able to use the system for 3, 4, or 5 trials each, all participants stated that they were interested in an at-home trial if one were made available, and no participants stopped any trial due to cybersickness. We did not find a statistically significant effect of trial on participants' self-reported average pain scores, but this is not surprising given that the study was not intended or powered to detect a clinical signal for efficacy.

Limitations

Because of the small number of participants, we cannot draw conclusions about efficacy nor do we have enough information to discuss generalizability. In addition, participants were not blinded to treatment, and the researcher working directly with the patients (AB) was also not blind to the purpose of the study.

Next Steps

Because in this study, participants only held the trackers in their *un*injured hand, we could not directly track improvement in movement on the injured side.

Although hand tracking was not available at the time this study was conducted, we have since updated the project to include markerless hand tracking, including finger tracking through the Oculus Quest. This can facilitate tracking of the injured hand, and may also improve the overall patient experience. Future work should examine whether markerless tracking and finger tracking improve patient outcomes.

Other potential easy modifications include changing the appearance of the avatar body, changing how the avatar limbs are controlled, and changing the appearance and placement of the targets. We have made this project open source with the goal of encouraging further exploration of deploying MVF in immersive VR using consumer systems.

This pilot study measured participant pain scores using a daily pain survey modified from the Brief Pain Inventory [18]. Future studies may use other standardized pain indicators, for example, Pain Self-Efficacy Questionnaire (PSEQ) [20], EuroQol-5D (EQ-5D) [21], or Pain Catastrophizing Scale (PCS) [22], as well as analyzing improved function as an indicator of positive outcome.

All 9 participants in our initial pilot study were interested in an at-home trial if one were made available. While at-home VR systems are still out of reach for most consumers, such systems can increase accessibility [23]. An at-home module would allow patients to use the module at will and could also be a useful channel of communication between the clinician and the patient. In such a system, safety and social interaction must be fully considered. Thus, any consent form for such an environment should reiterate potential risks and protection of patients' data must also be assured [24].

We note that this software is not a commercial system or patient-ready application. Rather, it is designed to be used as a potential starting point for research for other clinicians interested in exploring potential therapeutic uses of immersive VR for visual feedback on movement. We look forward to the contributions of other researchers in this area.

Conflicts of Interest

None declared.

Multimedia Appendix 1
All study materials, including consent form, interview protocol and survey questions.
[PDF File (Adobe PDF File), 511 KB - jmir_v23i5e16536_app1.pdf]

Multimedia Appendix 2
A video of two research assistants using the system to simulate the patient experience.

XSL•FO
RenderX

[MP4 File (MP4 Video), 72244 KB - jmir_v23i5e16536_app2.mp4]

Multimedia Appendix 3
Anonymized data set containing participants' self-reported pain ratings before and after the virtual reality sessions.
[XLSX File (Microsoft Excel File), 116 KB - jmir_v23i5e16536_app3.xlsx]

Multimedia Appendix 4
A representative example of participant movement data.
[PNG File , 23 KB - jmir_v23i5e16536_app4.png]

References

1. Rockett M. Diagnosis, mechanisms and treatment of complex regional pain syndrome. Curr Opin Anaesthesiol 2014 Oct;27(5):494-500 [FREE Full text] [doi: 10.1097/ACO.0000000000000114] [Medline: 25111604]
2. Di Lazzaro V, Pilato F, Dileone M, Profice P, Oliviero A, Mazzone P, et al. The physiological basis of the effects of intermittent theta burst stimulation of the human motor cortex. J Physiol 2008 Aug 15;586(16):3871-3879 [FREE Full text] [doi: 10.1113/jphysiol.2008.152736] [Medline: 18566003]
3. George MS, Taylor JJ, Short EB. The expanding evidence base for rTMS treatment of depression. Curr Opin Psychiatry 2013 Jan;26(1):13-18 [FREE Full text] [doi: 10.1097/YCO.0b013e32835ab46d] [Medline: 23154644]
4. Harden RN, Maihofner C, Abousaad E, Vatine J, Kirsling A, Perez RSGM, et al. A prospective, multisite, international validation of the Complex Regional Pain Syndrome Severity Score. Pain 2017 Aug;158(8):1430-1436. [doi: 10.1097/j.pain.0000000000000927] [Medline: 28715350]
5. Barad MJ, Ueno T, Younger J, Chatterjee N, Mackey S. Complex regional pain syndrome is associated with structural abnormalities in pain-related regions of the human brain. J Pain 2014 Feb;15(2):197-203 [FREE Full text] [doi: 10.1016/j.jpain.2013.10.011] [Medline: 24212070]
6. Gaertner M, Kong J, Scherrer KH, Foote A, Mackey S, Johnson KA. Advancing Transcranial Magnetic Stimulation Methods for Complex Regional Pain Syndrome: An Open-Label Study of Paired Theta Burst and High-Frequency Stimulation. Neuromodulation 2018 Jun;21(4):409-416 [FREE Full text] [doi: 10.1111/ner.12760] [Medline: 29504190]
7. Ramachandran VS, Altschuler EL. The use of visual feedback, in particular mirror visual feedback, in restoring brain function. Brain 2009 Jul;132(Pt 7):1693-1710. [doi: 10.1093/brain/awp135] [Medline: 19506071]
8. McCabe CS, Haigh RC, Ring EFJ, Halligan PW, Wall PD, Blake DR. A controlled pilot study of the utility of mirror visual feedback in the treatment of complex regional pain syndrome (type 1). Rheumatology (Oxford) 2003 Jan;42(1):97-101. [doi: 10.1093/rheumatology/keg041] [Medline: 12509620]
9. Sato K, Fukumori S, Matsusaki T, Maruo T, Ishikawa S, Nishie H, et al. Nonimmersive virtual reality mirror visual feedback therapy and its application for the treatment of complex regional pain syndrome: an open-label pilot study. Pain Med 2010 Apr;11(4):622-629. [doi: 10.1111/j.1526-4637.2010.00819.x] [Medline: 20202141]
10. Griffin A, Wilson L, Feinstein AB, Bortz A, Heirich MS, Gilkerson R, et al. Virtual Reality in Pain Rehabilitation for Youth With Chronic Pain: Pilot Feasibility Study. JMIR Rehabil Assist Technol 2020 Nov 23;7(2):e22620 [FREE Full text] [doi: 10.2196/22620] [Medline: 33226346]
11. Won AS, Bailey J, Bailenson J, Tataru C, Yoon IA, Golianu B. Immersive Virtual Reality for Pediatric Pain. Children (Basel) 2017 Jun 23;4(7):1-15 [FREE Full text] [doi: 10.3390/children4070052] [Medline: 28644422]
12. Won AS, Tataru CA, Cojocaru CM, Krane EJ, Bailenson JN, Niswonger S, et al. Two Virtual Reality Pilot Studies for the Treatment of Pediatric CRPS. Pain Med 2015 Aug;16(8):1644-1647. [doi: 10.1111/pme.12755] [Medline: 25930099]
13. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-Administered Skills-Based Virtual Reality Intervention for Chronic Pain: Randomized Controlled Pilot Study. JMIR Form Res 2020 Jul 07;4(7):e17293 [FREE Full text] [doi: 10.2196/17293] [Medline: 32374272]
14. Oculus. URL: https://www.oculus.com/ [accessed 2021-04-10]
15. VIVE. URL: https://www.vive.com/us/ [accessed 2021-04-10]
16. Center for Open Science. OSF. 2021. URL: https://osf.io/6pkns/files/ [accessed 2021-04-10]
17. Amtmann D, Cook KF, Jensen MP, Chen W, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182 [FREE Full text] [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]
18. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994 Mar;23(2):129-138. [Medline: 8080219]
19. Witmer BG, Singer MJ. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
20. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain 2007 Feb;11(2):153-163. [doi: 10.1016/j.ejpain.2005.12.008] [Medline: 16446108]
21. Balestroni G, Bertolotti G. [EuroQol-5D (EQ-5D): an instrument for measuring quality of life]. Monaldi Arch Chest Dis 2012 Sep;78(3):155-159 [FREE Full text] [doi: 10.4081/monaldi.2012.121] [Medline: 23614330]

XSL•FO
RenderX

22. Darnall BD, Sturgeon JA, Cook KF, Taub CJ, Roy A, Burns JW, et al. Development and Validation of a Daily Pain Catastrophizing Scale. J Pain 2017 Sep;18(9):1139-1149 [FREE Full text] [doi: 10.1016/j.jpain.2017.05.003] [Medline: 28528981]
23. Garrett B, Taverner T, McDade P. Virtual Reality as an Adjunct Home Therapy in Chronic Pain Management: An Exploratory Study. JMIR Med Inform 2017 May 11;5(2):e11 [FREE Full text] [doi: 10.2196/medinform.7271] [Medline: 28495661]
24. Bailenson J. Protecting Nonverbal Data Tracked in Virtual Reality. JAMA Pediatr 2018 Oct 01;172(10):905-906. [doi: 10.1001/jamapediatrics.2018.1909] [Medline: 30083770]

Abbreviations

CRPS: complex regional pain syndrome
EQ-5D: EuroQol-5D
MVF: mirror visual feedback
PCS: Pain Catastrophizing Scale
PSEQ: Pain Self-Efficacy Questionnaire
VR: virtual reality

Edited by R Kukafka; submitted 03.11.19; peer-reviewed by M Linnett, D Tu; comments to author 26.10.20; revised version received 07.01.21; accepted 16.03.21; published 26.05.21.

Please cite as:
Won AS, Barreau AC, Gaertner M, Stone T, Zhu J, Wang CY, Mackey S
Assessing the Feasibility of an Open-Source Virtual Reality Mirror Visual Feedback Module for Complex Regional Pain Syndrome: Pilot Usability Study
J Med Internet Res 2021;23(5):e16536
URL: https://www.jmir.org/2021/5/e16536
doi:10.2196/16536
PMID:34037530

©Andrea Stevenson Won, Ariana C Barreau, Mark Gaertner, Tristan Stone, Joshua Zhu, Cheng Yao Wang, Sean Mackey. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2021. 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

Virtual Reality for Sedation During Atrial Fibrillation Ablation in Clinical Practice: Observational Study

Thomas Roxburgh[1], MD; Anthony Li[2], MD; Charles Guenancia[3], MD, PhD; Patrice Pernollet[1], MD; Claire Bouleti[1], MD, PhD; Benjamin Alos[1], MD, MSc; Matthieu Gras[1], MD; Thomas Kerforne[1], MD, PhD; Denis Frasca[1], MD, PhD; François Le Gal[1], MD, PhD; Luc Christiaens[1], MD; Bruno Degand[1], MD; Rodrigue Garcia[1], MD, PhD

[1]Department of Anesthesia and Critical Care, University Hospital of Poitiers, Poitiers, France

[2]Cardiology Clinical Academic Group, St George's, University of London, London, United Kingdom

[3]Cardiology Department, University Hospital of Dijon, Dijon, France

Corresponding Author:
Rodrigue Garcia, MD, PhD
Department of Anesthesia and Critical Care
University Hospital of Poitiers
2 Rue de la Milétrie
Poitiers, 86021
France
Phone: 33 549443729
Email: rodrigue.garcia@chu-poitiers.fr

Abstract

Background: Connected devices are dramatically changing many aspects in health care. One such device, the virtual reality (VR) headset, has recently been shown to improve analgesia in a small sample of patients undergoing transcatheter aortic valve implantation.

Objective: We aimed to investigate the feasibility and effectiveness of VR in patients undergoing atrial fibrillation (AF) ablation under conscious sedation.

Methods: All patients who underwent an AF ablation with VR from March to May 2020 were included. Patients were compared to a consecutive cohort of patients who underwent AF ablation in the 3 months prior to the study. Primary efficacy was assessed by using a visual analog scale, summarizing the overall pain experienced during the ablation.

Results: The AF cryoablation procedure with VR was performed for 48 patients (mean age 63.0, SD 10.9 years; n=16, 33.3% females). No patient refused to use the device, although 14.6% (n=7) terminated the VR session prematurely. Preparation of the VR headset took on average 78 (SD 13) seconds. Compared to the control group, the mean perceived pain, assessed with the visual analog scale, was lower in the VR group (3.5 [SD 1.5] vs 4.3 [SD 1.6]; *P*=.004), and comfort was higher in the VR group (7.5 [SD 1.6] vs 6.8 [SD 1.7]; *P*=.03). On the other hand, morphine consumption was not different between the groups. Lastly, complications, as well as procedure and fluoroscopy duration, were not different between the two groups.

Conclusions: We found that VR was associated with a reduction in the perception of pain in patients undergoing AF ablation under conscious sedation. Our findings demonstrate that VR can be easily incorporated into the standard ablation workflow.

(J Med Internet Res 2021;23(5):e26349) doi:10.2196/26349

KEYWORDS

connected devices; virtual reality; atrial fibrillation; pain management

Introduction

Atrial fibrillation (AF) ablation is an established therapy for patients with symptomatic AF [1]. The number of AF ablations is increasing year on year, but the availability of general anesthesia does not meet demand [2]. As such, it is now commonplace for procedures to be performed under conscious sedation. However, patients' pain and discomfort during AF ablation may be associated with poorer outcomes [3].

Connected devices are dramatically changing many aspects in health care [4-7]. While in cardiology, the majority of devices are intended to monitor heart rhythm, others have a therapeutic

XSL•FO
RenderX

purpose [8-10]. One such device, the virtual reality headset, is the subject of numerous studies [11-15] and has recently shown to improve analgesia in a small sample of patients undergoing transcatheter aortic valve implantation [16,17]. To our knowledge, this device has not been tested in other cardiac procedures.

We aimed to investigate the feasibility and effectiveness of VR in patients undergoing catheter ablation of AF under conscious sedation.

Methods

Recruitment

From March to May 2020, all consecutive patients in whom an AF ablation was performed using cryoballoon (Arctic Front Advance; Medtronic) at the University Hospital of Poitiers were included. All participants received VR using a Deepsen headset on top of the usual analgesia protocol. The VR technique uses cognitive saturation in association with cardiac coherence breathing, music therapy, and gamification [18]. At the beginning of the session, 5 minutes of cardiac coherence breathing was delivered. Then, the patient was immersed in 1 of 5 3D computer-simulated scenarios. During the whole procedure, music therapy and gamification were also used. The patient played an active role as he or she interacted with the virtual environment, which aimed to deepen the immersion to unconsciously disconnect the patient from painful moments in the procedure. Our analgesia protocol consisted of 1 g of intravenous (IV) paracetamol, 20 mg of IV nefopam, 1 mg of IV midazolam, and 3 mg of IV morphine just before the start of the procedure. Patients could also request additional analgesia, in which case further boluses of 1 mg of morphine were given. Patients who underwent AF ablation with VR were compared to a consecutive cohort of patients who received routine AF cryoablation in the 3 months prior to the study using the standard analgesia protocol.

Outcomes

The feasibility of VR was assessed by the number of patients who refused this technique, the tolerance of the VR headset, and the time taken to install the device. Primary efficacy was assessed by using a visual analog scale (VAS), shown to the patient 45 minutes post procedure when they were asked to select a single point on the scale to summarize the overall pain experienced during the ablation [19]. The maximum pain intensity perceived was also recorded using a VAS score. Finally, the patient's comfort was assessed using a numerical scale, with 0 being "the most uncomfortable procedure you could have" and 10 being "the most comfortable procedure you could have." Oral informed consent was obtained from all participants. According to French legislation, this study was declared to the Commission Informatique et Libertés and did not require the approval of an Ethics Committee as this device is CE marked and is already used in routine clinical practice in some centers in France.

Statistical Analysis

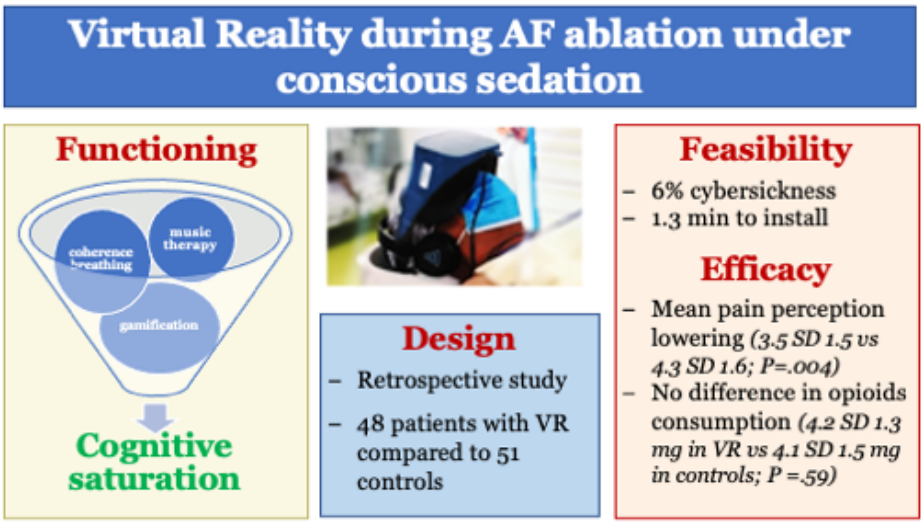
Continuous variables were expressed as mean (SD) and categorical variables were presented as numbers and percentages. Comparisons between groups were performed using the Student *t* test or the Mann-Whitney *U* test for continuous variables as appropriate, and the chi-square test for categorical variables. Analyses were performed using SPSS 22 (IBM Corp) statistical software.

Results

Feasibility

A total of 48 patients were enrolled to receive VR during AF cryoablation procedure (mean age 63.0, SD 10.9 years; n=16, 33.3% females). No patient refused to use the device, although 7 (14.6%) terminated the VR session prematurely. Four patients had a vasovagal reaction and 3 experienced cybersickness (vertigo: n=2; headache: n=1) (Figure 1). Preparation of the VR headset took on average 78 (SD 13) seconds.

Figure 1. Virtual reality (VR) use during atrial fibrillation (AF) ablation.



XSL•FO
RenderX

Efficacy

The control group comprised 51 patients. No significant differences were found between control and intervention groups in terms of age, gender, New York Heart Association functional class, left ventricular ejection fraction, or previous medications (Table 1).

Table 1. Baseline characteristics.

Characteristic	Virtual reality group (n=48)	Control group (n=51)	*P* value
Age (years), mean (SD)	63.0 (10.9)	64.5 (10.4)	.46
Gender (male), n (%)	32 (67)	39 (76)	.28
BMI (kg/m^2), mean (SD)	28.0 (4.6)	27.4 (4.9)	.55
NYHA functional class[a], n (%)			.15
Class I-II	30 (83)	33 (89)	
Class III-IV	6 (17)	4 (11)	
Hypertension, n (%)	19 (40)	17 (53)	.18
Diabetes mellitus, n (%)	5 (10)	2(3)	.21
Atrial fibrillation, n (%)			.23
Paroxysmal	19 (40)	27 (53)	
Persistent	29 (60)	24 (47)	
Long-standing persistent	0 (0)	0 (0)	
CHA_2DS_2-VASc[b] score (SD)			.66
0-1	15 (32)	15 (29)	
2-3	29 (60)	29 (57)	
≥4	4 (8)	7 (14)	
Ischemic cardiomyopathy, n (%)	4 (8)	11 (22)	.06
Systolic blood pressure (mmHg), mean (SD)	137 (18)	136 (23)	.77
Electrocardiogram			
Heart rate (bpm), mean (SD)	72 (19)	74 (22)	.63
Sinus rhythm, n (%)	30 (63)	35 (69)	.79
Echocardiography, mean (SD)			
Left ventricular ejection fraction (%)	59 (9)	59 (10)	.74
Left atrial volume (mL/m^2)	76 (28)	70 (23)	.18
NT-proBNP[c] (ng/L), mean (SD)	656 (956)	760 (1127)	.25
Medication, n (%)			
Anticoagulant	47 (98)	50 (98)	.72
Beta-blockers	44 (92)	46 (90)	.80
Angiotensin-converting enzyme inhibitor	23 (48)	22 (43)	.63
Antiplatelet agents	2 (4)	5 (10)	.23
Amiodarone	19 (40)	29 (57)	.36

[a]A New York Heart Association (NYHA) functional class was calculated for patients with heart failure (n=36 in the virtual reality group and n=37 in the control group).

[b]CHA_2DS_2-VASc: congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65-74 years, sex category.

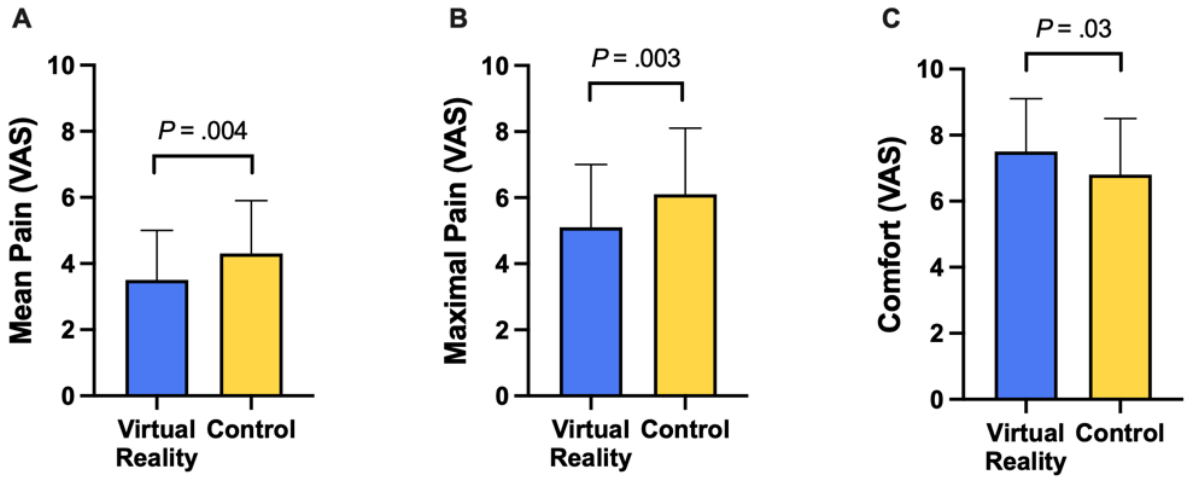
[c]NT-proBNP: N-terminal pro B-type natriuretic peptide.

Mean and maximal perceived pain were lower in the VR group (mean pain: 3.5 [SD 1.5] vs 4.3 [SD 1.6]; *P*=.004; maximal pain: 5.1 [SD 1.9] vs 6.1 [SD 2.0]; *P*=.003) and comfort was higher in the VR group (7.5 [SD 1.6] vs 6.8 [SD 1.7]; *P*=.03) (Figure 2). On the other hand, morphine consumption was not different across the groups (VR: 4.2 [SD 1.3] mg vs control:

XSL•FO
RenderX

4.1 [SD 1.5] mg; *P*=.59). In addition, procedure and fluoroscopy duration were not different between the two groups. Lastly, 3 (6.3%) patients had transient phrenic palsy in the VR group vs 6 (11.8%) in the control group (*P*=.34) and vagal reaction occurred in 4 (8.3%) patients in the VR group vs 2 (3.9%) in the control group (*P*=.60).

Figure 2. Mean pain, maximal pain, and comfort assessment in the virtual reality and control groups.



Discussion

Principal Findings

To our knowledge, this study is the first to evaluate the use of immersive VR to improve pain perception during cryoballoon AF ablation. Less than 10% of patients experienced cybersickness, and the device was set up in under 2 minutes. Compared to the control group, VR during AF ablation was associated with pain reduction and comfort improvement but did not lead to a reduction in opioid consumption.

Effect of VR During the AF Ablation Procedure

Several studies have assessed the feasibility and/or effectiveness of VR in different clinical settings [20-22]. However, the number of patients included in these studies was small. The present study showed that in patients undergoing AF ablation, VR uptake was high and well tolerated, suggesting that it can be widely adopted during AF ablation procedures performed under conscious sedation. Moreover, setup can be carried out by the operator or the nurse, and increased the total duration of the procedure by less than 2% [23], which is acceptable, as the device was associated with lower pain with the same level of opioid consumption. It can be speculated whether the absence of a placebo in the control group could have influenced the results. However, we could argue that the placebo effect of VR is still valid because it is really this subjective assessment of pain reduction and comfort increase that are the endpoints we are trying to achieve. Nevertheless, the reduction in pain perception seen in this study did not result in reduced opioid use since morphine consumption was not different between the two groups. This lack of difference might be related to the low level of additional opioids required after the initial bolus of morphine. Indeed, only a mean of 1.2 mg and 1.1 mg were needed on top of the initial bolus in the VR and control groups, respectively. Overall, our results are in line with the literature [24]. In a recent systematic review, Smith et al [24] suggested that VR was effective for analgesia in a variety of different clinical settings but could also have disadvantages.

Advantages and Drawbacks of VR

Other alternative techniques such as music therapy and hypnosis have been developed to relieve pain during conscious sedation anesthesia, but each have their drawbacks and advantages [25,26]. As the brain can only process a limited amount of information, mind saturation using VR aims to increase nonpainful input and limit the transmission of pain information according to the gate control theory [27]. Moreover, contrary to hypnosis, VR has a minimal learning curve and does not require specialist training, which may facilitate widespread adoption by other health care centers [19]. On the other hand, VR may elicit unpleasant reactions, such as cybersickness, in patients prone to vertigo or seasickness caused by conflicting sensory signals [28]. While the patient receives visual signals informing him or her that he or she is moving, no corroborating information is provided by the vestibular organs. Cybersickness has been described to occur in 20% to 80% of cases [29], although it occurred in only 18.8% of the sample in Bruno et al's [17] study. In our study, the occurrence of this side effect was even lower (6.3%) and may be due to younger age, fewer comorbidities, and better hemodynamic stability in patients undergoing AF ablation compared to those who underwent the transcatheter aortic valve implantation procedure.

Limitations

This study was a nonrandomized, single-center study. Nevertheless, the relatively high number of patients in each group and consecutive inclusion have limited bias. Moreover, baseline characteristics were not different between groups. Finally, patients' prior experience with interactive games may have influenced the effect of VR and its associated side effects, which was not systematically assessed or taken into account in our study.

XSL•FO
RenderX

Conclusion

Our study demonstrates that VR can be easily incorporated into the standard AF ablation workflow. Further, it was associated with a reduction in the perception of pain, even if it did not result in less opioid consumption, and improved patient experience. Larger randomized studies are needed to confirm these promising findings.

Acknowledgments

We deeply thank Laurent Cornu, David Czerwinski, Laurane Nicolleau, Camille Delmotte, and Fanny Bonnefous for contributing to the study.

Conflicts of Interest

None declared.

References

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
2. Gandjbakhch E, Mandel F, Dagher Y, Hidden-Lucet F, Rollin A, Maury P. Incidence, epidemiology, diagnosis and prognosis of atrio-oesophageal fistula following percutaneous catheter ablation: a French nationwide survey. Europace 2021 Apr 06;23(4):557-564. [doi: 10.1093/europace/euaa278] [Medline: 33221901]
3. Wasserlauf J, Kaplan RM, Walega DR, Arora R, Chicos AB, Kim SS, et al. Patient-reported outcomes after cryoballoon ablation are equivalent between moderate sedation and general anesthesia. J Cardiovasc Electrophysiol 2020 Jul;31(7):1579-1584. [doi: 10.1111/jce.14547] [Medline: 32400079]
4. Chen SC, Hu R, McAdam R. Smart, Remote, and Targeted Health Care Facilitation Through Connected Health: Qualitative Study. J Med Internet Res 2020 Apr 28;22(4):e14201 [FREE Full text] [doi: 10.2196/14201] [Medline: 32343254]
5. Yen Y, Tsai Y, Su VY, Chan S, Yu W, Ho H, et al. Use and Cost-Effectiveness of a Telehealth Service at a Centralized COVID-19 Quarantine Center in Taiwan: Cohort Study. J Med Internet Res 2020 Dec 11;22(12):e22703 [FREE Full text] [doi: 10.2196/22703] [Medline: 33259324]
6. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of Telehealth During the COVID-19 Pandemic: Scoping Review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
7. Signal V, McLeod M, Stanley J, Stairmand J, Sukumaran N, Thompson D, et al. A Mobile- and Web-Based Health Intervention Program for Diabetes and Prediabetes Self-Management (BetaMe/Melon): Process Evaluation Following a Randomized Controlled Trial. J Med Internet Res 2020 Dec 01;22(12):e19150 [FREE Full text] [doi: 10.2196/19150] [Medline: 33258776]
8. Manninger M, Kosiuk J, Zweiker D, Njeim M, Antolic B, Kircanski B, et al. Role of wearable rhythm recordings in clinical decision making-The wEHRAbles project. Clin Cardiol 2020 Sep 22;43(9):1032-1039 [FREE Full text] [doi: 10.1002/clc.23404] [Medline: 32700414]
9. Kwon S, Hong J, Choi E, Lee B, Baik C, Lee E, et al. Detection of Atrial Fibrillation Using a Ring-Type Wearable Device (CardioTracker) and Deep Learning Analysis of Photoplethysmography Signals: Prospective Observational Proof-of-Concept Study. J Med Internet Res 2020 May 21;22(5):e16443 [FREE Full text] [doi: 10.2196/16443] [Medline: 32348254]
10. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
11. Tashjian VC, Mosadeghi S, Howard AR, Lopez M, Dupuy T, Reid M, et al. Virtual Reality for Management of Pain in Hospitalized Patients: Results of a Controlled Trial. JMIR Ment Health 2017 Mar 29;4(1):e9 [FREE Full text] [doi: 10.2196/mental.7387] [Medline: 28356241]
12. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PLoS One 2019;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
13. Faruki A, Nguyen T, Proeschel S, Levy N, Yu J, Ip V, et al. Virtual reality as an adjunct to anesthesia in the operating room. Trials 2019 Dec 27;20(1):782 [FREE Full text] [doi: 10.1186/s13063-019-3922-2] [Medline: 31882015]
14. Freitas DMDO, Spadoni VS. Is virtual reality useful for pain management in patients who undergo medical procedures? Einstein (Sao Paulo) 2019 May 20;17(2):eMD4837 [FREE Full text] [doi: 10.31744/einstein_journal/2019MD4837] [Medline: 31116237]
15. Brown L, Chen ET, Binder DS. The use of virtual reality for Peri-procedural pain and anxiety at an outpatient spine clinic injection visit: an exploratory controlled randomized trial. Am J Transl Res 2020;12(9):5818-5826 [FREE Full text] [Medline: 33042461]

XSL•FO
RenderX

16. Lier EJ, Oosterman JM, Assmann R, de Vries M, van Goor H. The effect of Virtual Reality on evoked potentials following painful electrical stimuli and subjective pain. Sci Rep 2020 Jun 03;10(1):9067 [FREE Full text] [doi: 10.1038/s41598-020-66035-4] [Medline: 32494060]
17. Bruno RR, Lin Y, Wolff G, Polzin A, Veulemans V, Klein K, et al. Virtual reality-assisted conscious sedation during transcatheter aortic valve implantation: a randomised pilot study. EuroIntervention 2020 Dec 18;16(12):e1014-e1020 [FREE Full text] [doi: 10.4244/EIJ-D-20-00269] [Medline: 32597390]
18. Morris BJ, Croker S, Zimmerman C, Gill D, Romig C. Gaming science: the "Gamification" of scientific thinking. Front Psychol 2013 Sep 09;4:607 [FREE Full text] [doi: 10.3389/fpsyg.2013.00607] [Medline: 24058354]
19. Garcia R, Bouleti C, Li A, Frasca D, El Harrouchi S, Marechal J, et al. Hypnosis Versus Placebo During Atrial Flutter Ablation: The PAINLESS Study: A Randomized Controlled Trial. JACC Clin Electrophysiol 2020 Nov;6(12):1551-1560. [doi: 10.1016/j.jacep.2020.05.028] [Medline: 33213815]
20. Atzori B, Lauro Grotto R, Giugni A, Calabrò M, Alhalabi W, Hoffman HG. Virtual Reality Analgesia for Pediatric Dental Patients. Front Psychol 2018 Nov 23;9:2265 [FREE Full text] [doi: 10.3389/fpsyg.2018.02265] [Medline: 30532720]
21. Chad R, Emaan S, Jillian O. Effect of virtual reality headset for pediatric fear and pain distraction during immunization. Pain Manag 2018 May;8(3):175-179. [doi: 10.2217/pmt-2017-0040] [Medline: 29722606]
22. Delion M, Klinger E, Bernard F, Aubin G, Minassian AT, Menei P. Immersing Patients in a Virtual Reality Environment for Brain Mapping During Awake Surgery: Safety Study. World Neurosurg 2020 Feb;134:e937-e943. [doi: 10.1016/j.wneu.2019.11.047] [Medline: 31734424]
23. Andrade JG, Champagne J, Dubuc M, Deyell MW, Verma A, Macle L, CIRCA-DOSE Study Investigators. Cryoballoon or Radiofrequency Ablation for Atrial Fibrillation Assessed by Continuous Monitoring: A Randomized Clinical Trial. Circulation 2019 Nov 26;140(22):1779-1788. [doi: 10.1161/CIRCULATIONAHA.119.042622] [Medline: 31630538]
24. Smith V, Warty RR, Sursas JA, Payne O, Nair A, Krishnan S, et al. The Effectiveness of Virtual Reality in Managing Acute Pain and Anxiety for Medical Inpatients: Systematic Review. J Med Internet Res 2020 Nov 02;22(11):e17980 [FREE Full text] [doi: 10.2196/17980] [Medline: 33136055]
25. Scaglione M, Battaglia A, Di Donna P, Peyracchia M, Bolzan B, Mazzucchi P, et al. Hypnotic communication for periprocedural analgesia during transcatheter ablation of atrial fibrillation. Int J Cardiol Heart Vasc 2019 Sep;24:100405 [FREE Full text] [doi: 10.1016/j.ijcha.2019.100405] [Medline: 31388561]
26. Ebrahimi R, Tan W. Role of Music for Conscious Sedation During Invasive Cardiac Catheterization. Am J Cardiol 2018 Sep 15;122(6):1095-1097. [doi: 10.1016/j.amjcard.2018.05.039] [Medline: 30072130]
27. Pereira PJS, Lerner EA. Gate Control Theory Springs a Leak. Neuron 2017 Feb 22;93(4):723-724 [FREE Full text] [doi: 10.1016/j.neuron.2017.02.016] [Medline: 28231458]
28. Weech S, Kenny S, Barnett-Cowan M. Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. Front Psychol 2019;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
29. Gallagher M, Dowsett R, Ferrè ER. Vection in virtual reality modulates vestibular-evoked myogenic potentials. Eur J Neurosci 2019 Nov;50(10):3557-3565. [doi: 10.1111/ejn.14499] [Medline: 31233640]

Abbreviations

AF: atrial fibrillation
IV: intravenous
VAS: visual analog scale
VR: virtual reality

Edited by R Kukafka; submitted 08.12.20; peer-reviewed by D Duncker, G Boriani, P Dilaveris; comments to author 15.03.21; revised version received 23.03.21; accepted 04.05.21; published 27.05.21.

Please cite as:
Roxburgh T, Li A, Guenancia C, Pernollet P, Bouleti C, Alos B, Gras M, Kerforne T, Frasca D, Le Gal F, Christiaens L, Degand B, Garcia R
Virtual Reality for Sedation During Atrial Fibrillation Ablation in Clinical Practice: Observational Study
J Med Internet Res 2021;23(5):e26349
URL: https://www.jmir.org/2021/5/e26349
doi:10.2196/26349
PMID:34042589

©Thomas Roxburgh, Anthony Li, Charles Guenancia, Patrice Pernollet, Claire Bouleti, Benjamin Alos, Matthieu Gras, Thomas Kerforne, Denis Frasca, François Le Gal, Luc Christiaens, Bruno Degand, Rodrigue Garcia. Originally published in the Journal

XSL•FO
RenderX

of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

Information Quality Frameworks for Digital Health Technologies: Systematic Review

Kayode Philip Fadahunsi[1], MBBS, MPH; Siobhan O'Connor[2], BSc, CIMA CBA, RN, FHEA, PhD; James Tosin Akinlua[3], MBBS, MPH, PhD; Petra A Wark[3,4], MSc, PhD, FHEA; Joseph Gallagher[5], BA, MD, MBChB, BAO, MRCPI, MICGP, FHFA; Christopher Carroll[6], BA, MA, MSc, PhD; Josip Car[1,7], FFPH, DIC, MSc, PhD, MD, FRCP; Azeem Majeed[1], MD, FRCP, FRCGP, FFPH; John O'Donoghue[8,9], BSc, MSc, PhD, Cert(ICL), FHEA, FBCS

[1]Department of Public Health and Primary Care, Imperial College London, London, United Kingdom

[2]School of Health in Social Science, The University of Edinburgh, Edinburgh, United Kingdom

[3]Department of Primary Care and Population Health Sciences, University College London, London, United Kingdom

[4]Centre for Intelligent Healthcare, Institute of Health and Wellbeing, Coventry University, Coventry, United Kingdom

[5]gHealth Research Group, School of Medicine, University College Dublin, Dublin, Ireland

[6]Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield, Sheffield, United Kingdom

[7]Centre for Population Health Sciences, LKC Medicine, Nayang Technological University, Sungapore, Singapore

[8]Malawi eHealth Research Centre, University College Cork, Cork, Ireland

[9]ASSERT Research Centre, University College Cork, Cork, Ireland

Corresponding Author:
Kayode Philip Fadahunsi, MBBS, MPH
Department of Public Health and Primary Care
Imperial College London
The Reynolds Building
St. Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 07477854209
Email: K.fadahunsi14@imperial.ac.uk

Abstract

Background: Digital health technologies (DHTs) generate a large volume of information used in health care for administrative, educational, research, and clinical purposes. The clinical use of digital information for diagnostic, therapeutic, and prognostic purposes has multiple patient safety problems, some of which result from poor information quality (IQ).

Objective: This systematic review aims to synthesize an IQ framework that could be used to evaluate the extent to which digital health information is fit for clinical purposes.

Methods: The review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines. We searched Embase, MEDLINE, PubMed, CINAHL, Maternity and Infant Care, PsycINFO, Global Health, ProQuest Dissertations and Theses Global, Scopus, and HMIC (the Health Management Information Consortium) from inception until October 2019. Multidimensional IQ frameworks for assessing DHTs used in the clinical context by health care professionals were included. A thematic synthesis approach was used to synthesize the Clinical Information Quality (CLIQ) framework for digital health.

Results: We identified 10 existing IQ frameworks from which we developed the CLIQ framework for digital health with 13 unique dimensions: accessibility, completeness, portability, security, timeliness, accuracy, interpretability, plausibility, provenance, relevance, conformance, consistency, and maintainability, which were categorized into 3 meaningful categories: availability, informativeness, and usability.

Conclusions: This systematic review highlights the importance of the IQ of DHTs and its relevance to patient safety. The CLIQ framework for digital health will be useful in evaluating and conceptualizing IQ issues associated with digital health, thus forestalling potential patient safety problems.

XSL•FO
RenderX

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42018097142; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=97142

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-024722

(J Med Internet Res 2021;23(5):e23479) doi:10.2196/23479

KEYWORDS

digital health; patient safety; information quality

Introduction

Background

Digital health—the use of digital technologies for health—is increasingly recognized as a major driver of quality in health care [1]. Digital health technologies (DHTs), such as telemedicine, electronic health records (EHRs), clinical decision support systems (CDSS), mobile health, computerized physician order entry, electronic prescribing systems, and web-based health services, can improve access and quality of health care services [2,3]. DHTs generate a copious amount of information used in health care for administrative, educational, research, and clinical purposes [4,5]. However, the clinical use of digital information for diagnostic, therapeutic, and prognostic purposes has multiple patient safety problems, including significant harms and death, some of which result from poor information quality (IQ) [6-9]. For instance, a patient in the United Kingdom experienced a life-threatening allergic reaction following a medication error because of inaccessible allergy information in the EHR [6].

IQ refers to the extent to which information is fit for a specific purpose [10,11]. IQ is multidimensional, with each dimension describing a unique aspect of information [10,12]. For example, accuracy describes the extent to which information is correct, and accessibility describes the extent to which information is easily obtainable [12]. Dimensions relating to a specific context are traditionally integrated into a framework for evaluating IQ within the context [10,11]. One IQ framework for EHRs [13] has 11 dimensions and 3 categories, as shown in Textbox 1. The framework depicts the relationship between the dimensions by categorizing statistically measurable dimensions as objectivity, security-related dimensions as integrity, and dimensions relating to the usefulness of information to intended users as utility [13].

Textbox 1. Dimensions and categories in an information quality framework for electronic health record.

Objectivity

- Accuracy
- Completeness
- Consistency
- Timeliness

Utility

- Provenance
- Interpretability
- Usability
- Relevance

Integrity

- Privacy
- Confidentiality
- Secure access

Research Problem and Objective

Currently, there is no consensus on the definition of IQ dimensions in the context of the use of digital health information for clinical purposes. There is a lack of consistency in the terminology and definition of dimensions in existing IQ frameworks, limiting a common understanding of IQ requirements for DHTs [14]. Although previous literature reviews have attempted to define the IQ dimensions of digital health information, they focused on the use of digital health information for administrative and research purposes [14,15]. Identifying and defining IQ dimensions in the context of the use of digital health information for clinical purposes is especially important considering the patient safety implications of poor IQ, as discussed earlier [6-8]. This study aims to use an evidence-based approach to integrate dimensions from existing IQ frameworks, thus promoting a common understanding of IQ requirements. In addition, safety concerns may discourage health

XSL•FO
RenderX

care professionals from adopting DHTs. Although many general practitioners in the United Kingdom would support the deployment of more DHTs in primary care, they are concerned about the safety of digital health information [16]. Thus, there is a need for a framework that can be used to evaluate the extent to which digital health information is suitable for clinical purposes. The aim of this systematic review is to identify and define dimensions within existing IQ frameworks for DHTs and synthesize an IQ framework that can be used to evaluate the extent to which digital health information is fit for clinical purposes, either diagnostic, therapeutic, or prognostic.

Methods

Review Checklist

The systematic review is reported based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist [17] presented in Multimedia Appendix 1.

Review Questions

The systematic review will address the following questions:

1. What IQ frameworks currently exist for evaluating DHTs?
2. How are dimensions within these existing IQ frameworks defined?
3. Which IQ dimensions indicate how well digital health information is fit for diagnostic, therapeutic, or prognostic purposes?
4. How are these digital health IQ dimensions related to one another?

Eligibility Criteria

The eligibility criteria of this review were based on a specific approach for identifying frameworks, theories, and models in a systematic review using behavior of phenomenon of interest, health context and model or theory [18,19]. The traditional population, intervention, comparator, and outcome approach was not suitable as we synthesized frameworks rather than interventions.

We included IQ frameworks for assessing DHTs used for clinical purposes but excluded frameworks for nonclinical or administrative purposes because they are less likely to affect patient safety. For example, an incidence reporting system within a hospital setting can be used for administrative purposes. Similarly, we excluded IQ frameworks for web-based health-related information and electronic learning because they are not directly used in the clinical management of patients at the point of care. We excluded self-management apps used by patients mainly for health education and disease tracking purposes, as their IQ requirements are probably different from those used for clinical purposes by health care professionals [20]. We included multidimensional frameworks, but not individual IQ dimensions, as IQ is an interrelated multidimensional concept. Both published and gray literature were included. The included studies were not restricted based on publication date, and all eligible studies until October 2019 were included. Restrictions based on publication status, study type, and publication date may inadvertently lead to the exclusion of potentially relevant IQ frameworks. A summary of the eligibility criteria is presented in Table 1.

Table 1. Inclusion and exclusion criteria.

Concept	Inclusion	Exclusion
Behavior of phenomenon of interest	Information quality or data quality	Information quality or data quality of administrative and nonclinical data
Health context	Use of digital health information for clinical purposes (ie, diagnostic, therapeutic, or prognostic)	Web-based search for health-related information, electronic learning, and digital health apps for self-management
Model or theory	Multidimensional framework	Individual dimension
Language	English	Non-English
Publication status	Published and gray literature	None
Date of publication	Any	None
Type of study	Any	None

Information Sources

We searched bibliographic health care databases, including Embase, MEDLINE, PubMed, CINAHL, Maternity and Infant Care, PsycINFO, and Global Health. We also searched Scopus to identify digital health publications in non–health care disciplines, such as engineering and computer science. In addition, we searched HMIC (the Health Management Information Consortium) and ProQuest Dissertations and Theses Global, which are regarded as good sources of gray literature [21,22]. We manually searched the references of the included studies and tracked their citations to identify other eligible studies using Scopus and Google Scholar.

Search Strategy

The search terms are related to 3 main concepts: (1) IQ (behavior of the phenomenon of interest), (2) digital health (health context), and (3) framework (model or theory) [18,19]. The search terms relating to each of these concepts were combined using the *OR* connector. The results of the 3 categories were then combined using the *AND* connector. A librarian was consulted for input on the search strategy. Medical Subject Headings and free-text terms were used. Truncation and adjacency searching were used to increase the sensitivity of the search, as appropriate. The search strategy is presented in Multimedia Appendix 2.

XSL•FO
RenderX

Data Management

We removed duplicates using Endnote Reference Management Software (Clarivate), and additional duplicates not identified by the Endnote function were removed manually. The deduplicated data were then imported into Covidence (Veritas Health Innovation Ltd), a review-management software program that operates in partnership with Cochrane Collaboration and allows multiple reviewers to work on study selection simultaneously and independently.

Study Selection

The eligible studies were identified in 2 stages: title and abstract screening and full-text review. Titles and abstracts of the studies were screened for eligibility by 2 independent reviewers (KPF and JTA) using the criteria outlined in Table 1. Conflicts were resolved by discussion between the 2 reviewers and adjudicated by a third independent reviewer (JOD) when necessary. The full-text review of all studies selected during the screening was independently conducted by 2 reviewers (KPF and SOC), with disagreement resolved as described previously.

Data Extraction

Overall, 2 reviewers (KPF and SOC) independently extracted data from each eligible study using a prepiloted Microsoft Excel data extraction form. Other reviewers (JOD, CC, PAW, JG, JC, and AM) reviewed the extracted data to ensure the accuracy and completeness of the data. We extracted the study details, including authors, year of publication, country, affiliation, study aim, study design, and publication status. We also extracted IQ framework–related data, including the method of framework development, method of framework validation (when available), type of DHT, IQ dimensions and their verbatim definition, categories of IQ dimensions (when available), and metrics of IQ dimension measurement (when available).

These data elements were defined as follows:

- IQ frameworks for DHTs: A systematic integration of IQ dimensions to evaluate health information technologies used in the diagnosis, treatment, and prognosis of patients.
- IQ dimensions within the frameworks in digital health: These are the evaluation criteria within the IQ frameworks that specify the extent to which health information technologies are fit for clinical use.
- Definition of IQ dimensions in digital health: A clear description of what aspect of information each dimension assesses.
- Categories of dimensions within IQ frameworks in digital health: IQ dimensions are often categorized to depict the relationship between IQ dimensions in an IQ framework.
- Metrics of measurement of IQ dimensions in digital health: How each IQ dimension is measured, for example, questionnaire and mathematical formulas.

Quality Assessment

We assessed the quality of the included studies using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies [23]. Selecting this tool was difficult, as the included papers comprised a range of methodologies, including ethnography study, literature review, practice brief, and framework development, with some of the papers not explicitly stating their methodology. Therefore, some of the questions on the checklist were not applicable. Scores were not assigned, as this was not recommended by the checklist [23]. Studies were not excluded based on quality assessment outcome, as this was unlikely to have any major impact on the ultimate definition of the dimensions and the resulting IQ framework. However, the assessment provided a general idea about the quality of the development processes of the existing IQ frameworks and, therefore, the strength of the evidence [24].

Data Synthesis

In this review, the IQ framework was developed using a thematic synthesis approach comprising 3 key stages: coding, descriptive synthesis, and analytical synthesis [25]. Although codes and descriptive themes were generated directly from the extracted definition of IQ dimensions, analytical themes were interpretations that went beyond the original data.

In the first stage, we coded the verbatim definitions of IQ dimensions extracted from the existing IQ frameworks in the included papers. Coding was done by identifying the unique concepts from each definition of the IQ dimension and highlighting them using the text highlight function of Microsoft Word (Microsoft).

Second, we categorized the codes based on their similarities and differences and created a descriptive theme to capture the meaning of each category. Each descriptive theme was defined based on the meaning of the original code from which it was created. The descriptive themes created were regarded as the IQ dimensions of the new IQ framework for digital health. Coding and descriptive synthesis were performed by 2 independent reviewers (KPF and JTA) with adjudication by a third independent reviewer (JOD).

Finally, we conceptualized analytical themes by considering the interrelationship between the descriptive themes (IQ dimensions) based on their definitions. The conceptualization of the analytical themes from the descriptive themes in thematic synthesis has been described as controversial because it is influenced by the insight and judgment of the reviewers [25]. This stage was quite challenging because of the subjective nature and varying perspectives of the reviewers. The following procedures were used to avoid bias and to achieve a consensus. The lead author (KPF) categorized the IQ dimensions without revealing his proposed categories to other reviewers. The other reviewers were then invited to categorize the IQ dimensions individually and email their suggested categories with rationale to the lead author without copying other members of the team. The reviewers were specifically asked to reflect on the suitability of digital health information for clinical purposes and its impact on patient safety while categorizing the IQ dimensions. Overall, 2 reviewers (KPF and JOD) then collated the inputs and carefully assigned a category to each of the dimensions based on the most popular suggestions considered along with the rationale. The framework was then shared with all the members of the team for further inputs and adaptation, if necessary.

Thus, a new digital health IQ framework was developed by synthesizing existing IQ frameworks for DHTs. The IQ

XSL•FO
RenderX

dimensions in the new framework are descriptive themes that were generated directly from the definition of IQ dimensions within existing frameworks, whereas the IQ categories were generated from the higher-order analytical synthesis of the descriptive themes.

Ethics

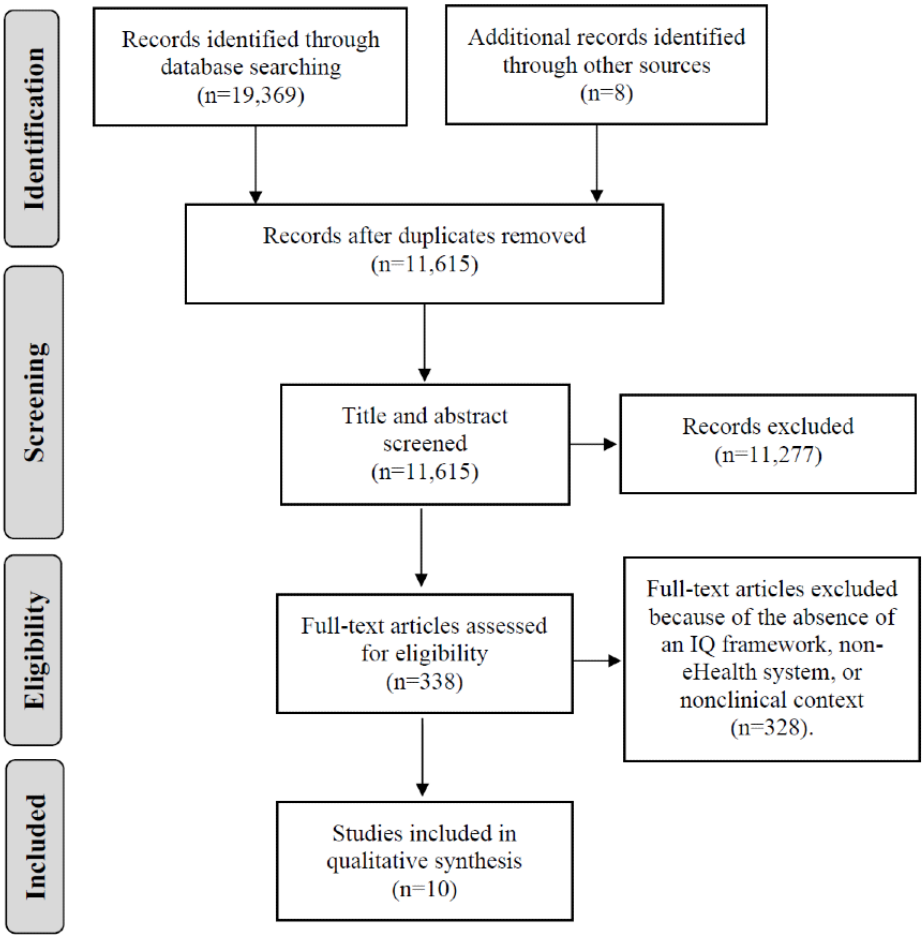
Ethical approval was not required for this systematic review, as the primary data were not collected. The review was registered in PROSPERO [26], and the protocol was published [27] to promote transparency.

Results

Selection of Studies

A total of 19,377 records were identified from the literature search. These were reduced to 338 after the removal of duplicates and screening of titles and abstracts. Only 10 papers were included in the study after a full-text review. Although 3 of these papers [14,28,29] were in the context of secondary use of digital health data for research, they were included, as their IQ frameworks were relevant to the clinical context of digital health information. However, we performed a sensitivity analysis by conducting a thematic synthesis with and without these 3 papers [30]. The sensitivity analysis revealed that the inclusion of the 3 papers did not affect the component dimensions in the resulting framework, but their inclusion produced a better understanding of the definition of the dimensions. The PRISMA flow diagram is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram. IQ: information quality.



Included Papers

The 10 included papers were published between 2007 and 2017. Of the 10 papers, 5 (50%) were published in the United States [14,28,31-33], 3 (30%) were published in the United Kingdom [13,29,34], and one each was published in Canada [35] and Japan [36]. One of the studies published in the United Kingdom was conducted in Saudi Arabia [13]. Of the 10 papers, 4 (40%) were journal publications [14,28,32,36], 3 (30%) were conference papers [29,31,34], 2 (20%) were institutional reports [33,35], and 1 (10%) was a PhD thesis [13]. Of the 10 studies, 5 used qualitative methods, either alone [31,36] or in combination with other methods [13,14,29]. Similarly, 40% (4/10) studies used literature review alone [28,34] or combined with other methods [13,14]. Overall, 30% (3/10) studies modified the existing frameworks [29,32,36]. One study reported to have updated the previous framework [33], but it was unclear how this was achieved. In addition, 10% (1/10) study [35] did not state how the framework was developed. About 50% (5/10) of the frameworks were on EHRs [13,14,28,32,33], one each

XSL•FO
RenderX

on electronic medical records [35], primary care databases [29], CDSSs [31], mobile and web-based apps for telemedicine [36], and cloud-based health information systems [34]. Thus, it appears that IQ framework research is unable to keep pace with the rapid evolution of DHTs with obvious underrepresentation of newer DHTs such as mobile health. The details of the included papers are presented in Multimedia Appendix 3 [13,14,28,29,31-36].

Quality of Included Studies

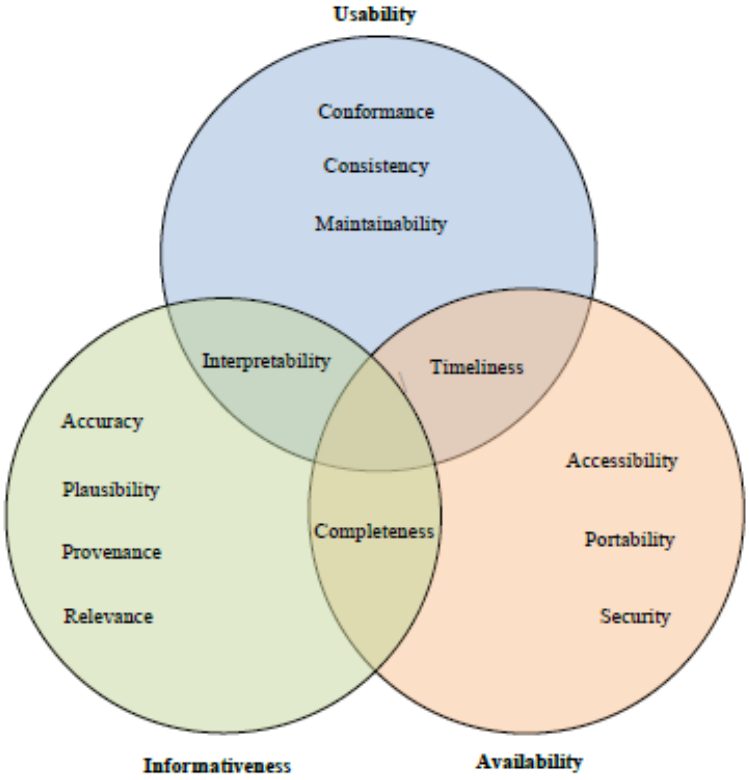
The quality assessment indicated that most of the studies described an IQ framework for DHT without reporting a robust framework development process. The Critical Appraisal Skills Programme checklist is not applicable to 2 studies [33,35], which are institutional publications. Only 1 qualitative study [13] reported on the recruitment strategy. Similarly, studies with literature reviews did not report on the search strategy or study selection process [13,14,28,34]. Only 3 studies [13,31,32] addressed ethical issues and reported sufficiently rigorous data analysis. These findings further justify the need for this study, which used a robust systematic review approach to develop a preliminary IQ framework for digital health. The quality assessment results are provided in Multimedia Appendix 4 [13,14,28,29,31-36].

Clinical Information Quality Framework for Digital Health

A total of 38 IQ dimensions and 70 verbatim definitions were extracted from the 10 included frameworks. The list of dimensions and their definitions are provided in Multimedia Appendix 5 [13,14,28,29,31-36]. The coding of these definitions led to the identification of 160 codes. Aggregation of similar codes resulted in a total of 13 unique IQ dimensions that mirrored all the relevant dimensions in the existing IQ frameworks while eliminating related but redundant dimensions. The resulting dimensions include accessibility, completeness, portability, security, timeliness, accuracy, interpretability, plausibility, provenance, relevance, conformance, consistency, and maintainability. These dimensions were defined based on the codes from which they were generated and classified into higher categories of availability, informativeness, and usability during the analytical synthesis. It is worth noting that some of the dimensions fit into more than one category but were placed into the best-fit category after carefully considering the inputs of all reviewers. For example, completeness was considered fit for both the availability and informativeness categories but was placed in the informativeness category, as this was the most popular category suggested by the reviewers. Similarly, timeliness was considered more fit for the availability category compared with the usability category, and interpretability was placed in the informativeness category rather than the usability category. The resulting Clinical Information Quality (CLIQ) framework for digital health is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Clinical Information Quality framework for digital health.



Informativeness directly concerns the usefulness of digital information for clinical purposes.	Accuracy	The extent to which information is correct.
	Completeness	The extent to which no required information is missing.
	Interpretability	The extent to which information can be understood.
	Plausibility	The extent to which information makes sense based on common knowledge.
	Provenance	The extent to which the source of information is trustworthy.
	Relevance	The extent to which information is useful for the intended task.
Availability concerns the functionality of the system holding clinical information.	Accessibility	The extent to which existing information is easily obtainable.
	Portability	The extent to which information is accessible in different systems.
	Security	The extent to which information is protected from unauthorized access and corruption.
	Timeliness	The extent to which current information is available on time.
Usability concerns the ease of use of clinical information.	Conformance	The extent to which information is presented in the desired format.
	Consistency	The extent to which information is presented in the same format.
	Maintainability	The extent to which information can be maintained.

Metrics of Measurement

Metrics of measurement for the IQ dimensions were given in only 30% (3/10) of the included papers [13,29,35]. The remaining papers only conceptualize IQ without providing guidance on its measurement. Objective and subjective measures were used in these studies. Objective measures involved mathematical calculations, such as ratio, percentages, and fraction to quantify the IQ dimension [37]. Subjective measures, on the other hand, rely on the perspectives of the information users, which are usually assessed using a Likert scale questionnaire or qualitative interviews [10].

Objective measures were reported for accuracy, validity, timeliness, completeness and interpretability, comprehensibility, reliability, validity, timeliness, relevance, integrity completeness, concordance, informative sufficiency, consistency, consistency of capture, and consistency of form [13,29,35]. These dimensions were measured by determining whether a desired

XSL•FO
RenderX

or undesired attribute was present or absent. For example, Almutiry [13] identified quality problems (undesired attributes) related to accuracy as misspelling, out-of-range values, erroneous values, etc. The quality score for accuracy was then calculated by determining the proportion of the total data units without each quality problem. Similarly, Dungey et al [29] measured accuracy by calculating the proportion of implausible values (undesired attributes). Bowen [35], on the other hand, used the percentage of data units with the desired attribute. Correctness was measured by determining the positive predictive value, which is the proportion of true positives (desired attributes).

Subjective measures were reported for usability, relevance, provenance, secure access, confidentiality, and privacy [13]. Each dimension was measured using multiple Likert scale questions. For example, relevance was assessed by the information users' rating of how far the information was relevant, useful, applicable, and appropriate to the task at hand [13]. The quality score for each dimension is the aggregate of all ratings for these different measures.

Discussion

Principal Findings

We identified 10 existing IQ frameworks for DHTs, and from these, we developed the CLIQ framework for digital health with 13 unique dimensions, including accessibility, completeness, portability, security, timeliness, accuracy, interpretability, plausibility, provenance, relevance, conformance, consistency, and maintainability, which were classified into 3 meaningful categories—availability, informativeness, and usability—based on our conceptualization of *fitness* of digital health information for clinical purposes.

The informativeness category directly concerns the usefulness of information for clinical purposes and has the greatest implications for patient safety. Problems with the dimensions in the category can directly lead to significant harm, as previously reported in the literature [6,8]. Accuracy is the most popular IQ dimension. However, this systematic review echoes the literature that IQ is not only about accuracy but also a multidimensional phenomenon [38]. Provenance and plausibility are unique IQ dimensions that can be regarded as proxies for accuracy, especially in situations where immediate and objective determination of accuracy is impractical. Provenance and plausibility can be easily determined subjectively. For example, knowing that the source of digital health information is a reputable institution (provenance) such as the World Health Organization would be reassuring, and an implausible value, such as a body temperature of 100°C (plausibility), would raise a serious concern. Interpretability is critical to the clinical use of digital information, as an incorrect interpretation may lead to significant harm. Hence, the inclusion of reference values with most laboratory results enhances the safe interpretation of the values.

The availability category of IQ dimensions concerns the functionality of a system that holds clinical information. These dimensions are critical as they can affect the efficiency of service delivery and are regarded as important by users of digital health information. Inaccessible digital information offers no real value to health professionals, as it cannot be used in the clinical management of patients. In addition, accessibility of clinical information wherever it is required (portability) and whenever it is required (timeliness) could be lifesaving, especially in emergency situations when the knowledge of a patient's medical history and current medications are essential. Timeliness, in the clinical context of digital health, also requires that health information is up to date. On the other hand, restriction of access to clinical information only to authorized users (security) protects the privacy and confidentiality of the patient and protects the information from corruption. Availability dimensions are illustrated by the UK's Summary Care Records [39], which contain up-to-date personal medical and medication history of patients and are accessible at the point of care (timeliness and accessibility) across different health care settings only to authorized health care professionals (security).

The usability category concerns the ease of use of health information. Consistency and conformance are akin to 2 sides of a coin, with consistency referring to the presentation of information in the same format within a system and conformance referring to the presentation of information in the desired format based on local guidelines or international standards. For example, it is important for an app to present blood glucose consistently using either gram per deciliter or millimoles per liter and conform with the recommended units in the local guidelines to avoid confusion, which may compromise patient safety. The last dimension in this category is maintainability. This refers to the extent to which the information can be maintained. Maintenance, in this context, covers a range of activities, including review, audit, update, and storage of clinical information to ensure that all other IQ requirements are met. For example, timeliness can be improved by updating the information in the DHTs, and accuracy can be improved through regular audits of the information generated by the DHTs.

Strength and Limitations

The main strength of our framework lies in the rigorous systematic review approach that was used to identify, define, and categorize IQ dimensions. In addition, our approach of synthesizing definitions rather than the traditional practice of simply cross-matching dimensions from different frameworks is more meaningful, as the definition expresses the real meaning of each dimension, and a dimension usually has heterogeneous definitions across different frameworks. In addition, focusing on the clinical context rather than the ever-changing DHTs, as in previous frameworks, we have developed a context-specific IQ framework that would be applicable or at least adaptable to a range of DHTs used in the clinical context, including novel ones that are currently underrepresented in IQ framework research. This approach differs from previous frameworks that focus on individual DHTs, such as EHR [13] and CDSS [31]. The consideration of the clinical purposes of DHTs is in consonance with the *fit-for-purpose* definition of IQ [10]. Moreover, the traditional practice of using the same clinical information across different DHTs (eg, the use of EHR information for CDSS) further justifies the need for a common IQ framework for the clinical context of DHTs.

XSL•FO
RenderX

However, the lack of information about the relative relevance of the IQ dimensions in the CLIQ framework and the optimal means of their measurement are limitations. Although these dimensions could be considered as indices of fitness of digital information for clinical purposes, we acknowledge the need to consult with clinical information users, such as doctors, nurses, and health service managers, as recommended in the literature [10,12]. Thus, the current CLIQ framework for digital health could be regarded as a preliminary framework to be tested in primary research studies. To build on this preliminary research, an international eDelphi study is currently underway to obtain consensus among clinicians on the approach to assessing the quality of clinical information produced by DHTs. The eDelphi study addresses the prioritization of the dimensions and the metrics for measuring the dimension.

Comparison With Validated IQ Frameworks

The CLIQ framework for digital health shares several characteristics with validated IQ frameworks within and beyond the health care domain. One such validated IQ framework, developed by Wang and Strong [38], has been used as a reference point in IQ research. Out of its 15 dimensions, 7 (accuracy, relevance, completeness, timeliness, interpretability, security, and accessibility) are also included in the CLIQ framework. The rest of its dimensions, such as believability and understandability, were assimilated by other dimensions in our framework during thematic synthesis. On the other hand, novel dimensions such as portability and maintainability are included in our framework but not in the framework developed by Wang and Strong [38]. This reflects technological advances in the last three decades, with an increasing amount of digital information. In addition, our framework was developed for the clinical context, whereas Wang and Strong focused on the business domain [38].

Similarly, the dimensions in our framework overlap with the product quality properties of the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC 25010), which include 8 characteristics: functional suitability, reliability, performance efficiency, operability, security, compatibility, maintainability, and transferability [40]. Of these characteristics, 2 (maintainability and security) were also included in the CLIQ framework. Other dimensions in the CLIQ framework (eg, availability, accuracy, and completeness) are included as subcharacteristics of product quality. This overlap is not unexpected, as DHTs are also software products, with IQ being a subset of product quality [40]. However, ISO/IEC 25010 addresses Systems and Software Quality Requirements and Evaluation from a computer engineering perspective, whereas the CLIQ framework addresses IQ from a health care perspective with consideration of its impact on patient safety [6-8]. Although we recognize the importance of other aspects of product quality, such as user-interface esthetics, these are beyond the scope of this study, which is focused on IQ in the clinical context of DHTs.

Conclusions

This systematic review highlighted the importance of the IQ of DHTs and their relevance to patient safety. Future research is needed to determine the relative relevance of each dimension in the CLIQ framework and their metrics of measurement, with inputs from clinical information users. The CLIQ framework for digital health will be useful to health care organizations, health care professionals, digital health solution developers, and medical device regulators in conceptualizing and evaluating IQ issues associated with digital health, thus forestalling potential patient safety problems. This is more relevant than ever, as the health care community is increasingly turning to DHTs, and the need for and value of such systems in the context of health emergencies is becoming ever more apparent.

Acknowledgments

The authors would like to thank Rebecca Jones, library manager and liaison librarian at the Charring Cross Library, Imperial College London, for useful advice on strategies for the literature search. This review is part of the PhD of the lead author (KPF) sponsored by the Federal Government of Nigeria. This review is an independent research supported in part by the National Institute for Health Research (NIHR) Applied Research Collaboration Northwest London. The views expressed in this publication 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

KPF and JOD conceived the study. KPF drafted the manuscript. JTA, SOC, PAW, CC, AM, JC, JG, and JOD revised the manuscript for important intellectual content and contributed to the review, including search strategy, study selection, data extraction, and data analysis. AM is the clinical lead, and JOD is the guarantor of the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[DOCX File , 18 KB - jmir_v23i5e23479_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Search strategy.
[DOCX File , 91 KB - jmir_v23i5e23479_app2.docx]

Multimedia Appendix 3
Included papers.
[DOCX File , 18 KB - jmir_v23i5e23479_app3.docx]

Multimedia Appendix 4
Quality assessment result.
[DOCX File , 15 KB - jmir_v23i5e23479_app4.docx]

Multimedia Appendix 5
Definition of dimensions within the existing information quality frameworks for digital health technologies.
[DOCX File , 48 KB - jmir_v23i5e23479_app5.docx]

References

1. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019:1-124.
2. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
3. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Saudi Med J 2017 Dec 03;38(12):1173-1180 [FREE Full text] [doi: 10.15537/smj.2017.12.20631] [Medline: 29209664]
4. World Health Organization. Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016:1-156.
5. Baur C, Kanaan SB. Expanding the reach and impact of consumer e-Health tools. US Department of Health and Human Services. Washington DC: US Department of Health and Human Services; 2006. URL: https://www.unapcict.org/resources/ictd-infobank/expanding-reach-and-impact-consumer-e-health-tools [accessed 2021-05-04]
6. Magrabi F, Baker M, Sinha I, Ong M, Harrison S, Kidd MR, et al. Clinical safety of England's national programme for IT: a retrospective analysis of all reported safety events 2005 to 2011. Int J Med Inform 2015 Mar;84(3):198-206. [doi: 10.1016/j.ijmedinf.2014.12.003]
7. Magrabi F, Ong M, Runciman W, Coiera E. Patient safety problems associated with heathcare information technology: an analysis of adverse events reported to the US Food and Drug Administration. AMIA Annu Symp Proc 2011;2011:853-857 [FREE Full text] [Medline: 22195143]
8. Magrabi F, Aarts J, Nohr C, Baker M, Harrison S, Pelayo S, et al. A comparative review of patient safety initiatives for national health information technology. Int J Med Inform 2013 May;82(5):139-148. [doi: 10.1016/j.ijmedinf.2012.11.014] [Medline: 23266061]
9. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. J Am Med Inform Assoc 2014 Nov 01;21(6):1053-1059 [FREE Full text] [doi: 10.1136/amiajnl-2013-002578] [Medline: 24951796]
10. Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. Inf Manag 2002 Dec;40(2):133-146. [doi: 10.1016/s0378-7206(02)00043-5]
11. Alkhattabi M, Neagu D, Cullen A. Information quality framework for e-learning systems. Knowl Manag E-Learning 2010;2(4):340-362. [doi: 10.34105/j.kmel.2010.02.025]
12. Eppler M, Wittig D. Conceptualizing information quality: a review of information quality frameworks from the last ten years. In: Proceedings of the Fifth Conference on Information Quality (IQ 2000). 2000 Presented at: Fifth Conference on Information Quality (IQ 2000); 2000; Massachusetts Institute of Technology, Cambridge, United States p. 83-96 URL: https://www.researchgate.net/publication/220918665_Conceptualizing_Information_Quality_A_Review_of_Information_Quality_Frameworks_from_the_Last_Ten_Years
13. Almutiry OS. Data quality assessment instrument for electronic health record systems in Saudi Arabia. University of Southampton, Doctoral Thesis: University of Southampton; 2017. URL: http://eprints.soton.ac.uk/id/eprint/419029 [accessed 2021-05-04]
14. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016 Sep 11;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]

XSL•FO
RenderX

15. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. Int J Environ Res Public Health 2014 May 14;11(5):5170-5207 [FREE Full text] [doi: 10.3390/ijerph110505170] [Medline: 24830450]
16. Iacobucci G. Digital health: GPs aren't "Luddites" but want safe, equitable care. Br Med J 2019;364:1258. [doi: 10.1136/bmj.l1258] [Medline: 30890525]
17. Moher D, Liberati A, Tetzlaff J, Altman D, 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]
18. Booth A, Carroll C. Systematic searching for theory to inform systematic reviews: is it feasible? Is it desirable? Health Info Libr J 2015 Sep 11;32(3):220-235 [FREE Full text] [doi: 10.1111/hir.12108] [Medline: 26095232]
19. Carroll C, Booth A, Leaviss J, Rick J. "Best fit" framework synthesis: refining the method. BMC Med Res Methodol 2013 Mar 13;13(1):37 [FREE Full text] [doi: 10.1186/1471-2288-13-37] [Medline: 23497061]
20. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
21. Haig A, Dozier M. BEME Guide no 3: systematic searching for evidence in medical education-Part 1: sources of information. Med Teach 2003 Jul 03;25(4):352-363. [doi: 10.1080/0142159031000136815] [Medline: 12893544]
22. Boukacem - Zeghmouri C, Schöpfel J. Document supply and open access: an international survey on grey literature. Interlend Doc Supply 2006 Jul;34(3):96-104. [doi: 10.1108/02641610610686012]
23. CASP Qualitative Checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2021-05-04]
24. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Version 1. ESRC Methods Programme 2006:-. [doi: 10.13140/2.1.1018.4643]
25. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8(1):45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
26. Fadahunsi K, Akinlua J, O'Connor S, Wark P, Gallagher J, Carroll C. Information quality frameworks in eHealth: a systematic review. PROSPERO CRD42018097142. 2018 May 23. URL: https://www.crd.york.ac.uk/PROSPEROFILES/97142_STRATEGY_20180521.pdf [accessed 2021-05-04]
27. Fadahunsi KP, Akinlua JT, O'Connor S, Wark PA, Gallagher J, Carroll C, et al. Protocol for a systematic review and qualitative synthesis of information quality frameworks in eHealth. BMJ Open 2019 Mar 05;9(3):e024722. [doi: 10.1136/bmjopen-2018-024722]
28. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
29. Dungey S, Beloff N, Puri S, Boggon R, Williams T, Tate A. A pragmatic approach for measuring data quality in primary care databases. In: Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 2014 Presented at: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); June 1-4, 2014; Valencia, Spain. [doi: 10.1109/BHI.2014.6864484]
30. Carroll C, Booth A, Lloyd-Jones M. Should we exclude inadequately reported studies from qualitative systematic reviews? An evaluation of sensitivity analyses in two case study reviews. Qual Health Res 2012 Oct 03;22(10):1425-1434. [doi: 10.1177/1049732312452937] [Medline: 22865107]
31. McCormack JL, Ash JS. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. AMIA Annu Symp Proc 2012;2012:1302-1309 [FREE Full text] [Medline: 23304409]
32. Stetson P, Bakken S, Wrenn J, Siegler E. Assessing electronic note quality using the Physician Documentation Quality Instrument (PDQI-9). Appl Clin Inform 2012 Dec 16;3(2):164-174 [FREE Full text] [doi: 10.4338/aci-2011-11-ra-0070] [Medline: 22577483]
33. Davoudi S, Dooling J, Glondys B, Jones T, Kadlec L, Overgaard S. Data quality management model (2015 update) - retired. American Health Information Management Association. 2015. URL: http://library.ahima.org/PB/DataQualityModel#.XW6r-pNKjab [accessed 2019-10-25]
34. Almutiry O, Wills G, Alwabel A, Crowder R, Walters R. Toward a framework for data quality in cloud-based health information system. In: Proceedings of the International Conference on Information Society (i-Society 2013). 2013 Presented at: International Conference on Information Society (i-Society 2013); June 24-26, 2013; Toronto, ON, Canada p. -.
35. Bowen M. EMR data quality evaluation guide. eHealth Observatory. 2012. URL: http://ehealth.uvic.ca/resources/tools/EMRsystem/2012.04.24-DataQualityEvaluationGuide-v1.0.pdf [accessed 2021-05-04]
36. Bolt T, Kano S, Kodate A. Information quality in home care coordination services. J Telemed Telecare 2016 Dec 02;13(1_suppl):7-9. [doi: 10.1258/135763307781645149]
37. Pipino LL, Lee YW, Wang RY. Data quality assessment. Commun ACM 2002 Apr;45(4):211-218. [doi: 10.1145/505248.506010]

XSL•FO
RenderX

38. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. J Manag Inform Syst 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
39. Summary Care Records (SCR). NHS Digital. 2020. URL: https://digital.nhs.uk/services/summary-care-records-scr [accessed 2021-05-04]
40. Systems and software engineering. Systems and software quality requirements and evaluation (SQuaRE). System and software quality models. BS ISO/IEC 25010:2011. 2011. URL: https://landingpage.bsigroup.com/LandingPage/Standard?UPI=000000000030215101 [accessed 2021-05-04]

Abbreviations

CDSS: clinical decision support system
CLIQ: Clinical Information Quality
DHT: digital health technology
EHR: electronic health record
IQ: information quality
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

Edited by G Eysenbach; submitted 13.08.20; peer-reviewed by D Strong, M Kolotylo-Kulkarni, A Sharafoddini, MDG Pimentel, A Azzam, C Jones, N Mohammad Gholi Mezerji; comments to author 18.11.20; revised version received 28.11.20; accepted 03.04.21; published 17.05.21.

Please cite as:
Fadahunsi KP, O'Connor S, Akinlua JT, Wark PA, Gallagher J, Carroll C, Car J, Majeed A, O'Donoghue J
Information Quality Frameworks for Digital Health Technologies: Systematic Review
J Med Internet Res 2021;23(5):e23479
URL: https://www.jmir.org/2021/5/e23479
doi:10.2196/23479
PMID:33835034

©Kayode Philip Fadahunsi, Siobhan O'Connor, James Tosin Akinlua, Petra A Wark, Joseph Gallagher, Christopher Carroll, Josip Car, Azeem Majeed, John O'Donoghue. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

Digitization of Measurement-Based Care Pathways in Mental Health Through REDCap and Electronic Health Record Integration: Development and Usability Study

Steve Hawley[1*], MSc, EMHI; Joanna Yu[1*], PhD; Nikola Bogetic[1], BSc; Natalia Potapova[1], BSc; Chris Wakefield[2]; Mike Thompson[2]; Stefan Kloiber[3], MD; Sean Hill[1], PhD; Damian Jankowicz[2], PhD; David Rotenberg[1], MSc

[1]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]Clinical Applications, Centre for Addiction and Mental Health, Toronto, ON, Canada

[3]General Adult Psychiatry and Health Systems Division, Centre for Addiction and Mental Health, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Steve Hawley, MSc, EMHI
Krembil Centre for Neuroinformatics
Centre for Addiction and Mental Health
250 College St.
Toronto, ON, M5T 1L8
Canada
Phone: 1 416 535 8501
Email: steve.hawley@camh.ca

Abstract

Background: The delivery of standardized self-report assessments is essential for measurement-based care in mental health. Paper-based methods of measurement-based care data collection may result in transcription errors, missing data, and other data quality issues when entered into patient electronic health records (EHRs).

Objective: This study aims to help address these issues by using a dedicated instance of REDCap (Research Electronic Data Capture; Vanderbilt University)—a free, widely used electronic data capture platform—that was established to enable the deployment of digitized self-assessments in clinical care pathways to inform clinical decision making.

Methods: REDCap was integrated with the primary clinical information system to facilitate the real-time transfer of discrete data and PDF reports from REDCap into the EHR. Both technical and administrative components were required for complete implementation. A technology acceptance survey was also administered to capture physicians' and clinicians' attitudes toward the new system.

Results: The integration of REDCap with the EHR transitioned clinical workflows from paper-based methods of data collection to electronic data collection. This resulted in significant time savings, improved data quality, and valuable real-time information delivery. The digitization of self-report assessments at each appointment contributed to the clinic-wide implementation of the major depressive disorder integrated care pathway. This digital transformation facilitated a 4-fold increase in the physician adoption of this integrated care pathway workflow and a 3-fold increase in patient enrollment, resulting in an overall significant increase in major depressive disorder integrated care pathway capacity. Physicians' and clinicians' attitudes were overall positive, with almost all respondents agreeing that the system was useful to their work.

Conclusions: REDCap provided an intuitive patient interface for collecting self-report measures and accessing results in real time to inform clinical decisions and an extensible backend for system integration. The approach scaled effectively and expanded to high-impact clinics throughout the hospital, allowing for the broad deployment of complex workflows and standardized assessments, which led to the accumulation of harmonized data across clinics and care pathways. REDCap is a flexible tool that can be effectively leveraged to facilitate the automatic transfer of self-report data to the EHR; however, thoughtful governance is required to complement the technical implementation to ensure that data standardization, data quality, patient safety, and privacy are maintained.

(J Med Internet Res 2021;23(5):e25656) doi:10.2196/25656

XSL•FO
RenderX

KEYWORDS

REDCap; electronic health record; systems integration; measurement-based care; hospital information systems

Introduction

Background

A key component of measurement-based care (MBC) in mental health relies on the ability of patients to complete self-assessments. These assessments can be used in conjunction with evidence-informed practice guidelines to monitor patient progress and inform clinical decisions for patient-centered care plans. MBC can improve outcomes by rapidly detecting patients with deteriorating symptoms and by effectively quantifying persistent symptoms [1]. The Centre for Addiction and Mental Health (CAMH) is the largest mental health hospital in Canada, supporting over 34,000 patients per year across diverse clinical programs. CAMH has developed and deployed a number of integrated care pathways (ICPs), which include MBC as the main, evidence-based aspect of care. Broadly used across many clinical domains, ICPs are structured care plans that follow evidence-based guidelines to improve patient outcomes [2] and may include definitions for specific medications or treatment strategies [3,4].

In addition to providing clinical care, CAMH is Canada's leading mental illness research facility. In recent years, REDCap (Research Electronic Data Capture; Vanderbilt University) has been instrumental to this research effort [5]. REDCap is a secure, Health Insurance Portability and Accountability Act–compliant, web-based application for building and managing electronic surveys and assessments [6]. Developed in 2004 by researchers at Vanderbilt University, REDCap was designed to provide a straightforward means of data management for the local research community [7]. Since its inception, the platform has evolved through several phases, first through select development partnerships and then through a growing international consortium that currently represents a diverse global community spanning over 3600 institutions in more than 130 countries with more than a million users worldwide [8]. The platform is available to academic, nonprofit, and government organizations through a no-cost consortium model, whereby partner institutions are provided the source code and installation files for REDCap after agreeing to the end user license agreements terms outlined by Vanderbilt University.

REDCap was launched at CAMH in 2015 for research data collection and operational support. Since then, the platform has grown to more than 1000 users who have created over 1800 projects culminating in more than 200,000 individual records. In 2017, a second, fully validated instance of REDCap was launched to support data collection for Health Canada regulated clinical trials. In 2018, because of the growing familiarity with REDCap at CAMH and knowledge of its flexibility, customizability, and community support, a demand for a version of REDCap to support clinical data began to emerge. Specifically, this version of REDCap would be used for ICP delivery. It was posited that REDCap's flexibility and its ability to manage complex workflows would significantly improve the previously paper-based methods used for ICP delivery. Importantly, however, this version of REDCap would need to be integrated with the CAMH electronic health record (EHR), a CAMH-branded version of Cerner Millennium called I-CARE.

CAMH achieved the Healthcare Information Management Systems Society (HIMSS) Adoption Model for Analytics Maturity Stage 6 in 2018 and the Electronic Medical Record Adoption Model Stage 7 in 2017 [9], indicating excellent hospital-wide adoption and compliance. Electronic data capture is used throughout the hospital and coordinated as the *sole source of truth* legal health records. To enhance the flexibility of electronic data collection for MBC, a *clinical* instance of REDCap was launched and connected to Cerner via a systems integration layer to facilitate the automatic transfer of completed assessments to the patient chart in both PDF and discrete formats. This integration extends capabilities for complex survey queues and branching assessments as needed by clinical pathways while maintaining the integrity of the health record.

Objectives

In this development and usability study, we describe the technical implementation and governance model of a clinical instance of REDCap to support digital MBC in multiple clinical environments. We detail the technical requirements for real-time transfer of self-assessments collected in REDCap to the patient's chart in Cerner in both PDF and discrete variable formats. In addition, we describe the clinical governance structure necessary to sustain high-quality clinical data and support standardized implementation across clinics. A pertinent clinical use case is outlined to illustrate the utility and applicability of REDCap to support scalable and extensible structured mental health MBC at CAMH and beyond.

Methods

Overview

Evaluation of the ICPs at CAMH identified a need to deploy complex bundles of assessments over time that could be dynamically adjusted based on indicators such as severity or frequency of symptoms. The functionality of REDCap is well suited to meet these requirements. By delivering standardized assessments longitudinally, modified through branching and survey queue logic as appropriate, necessary psychological dimensions could be sampled to inform physician decision making throughout the course of care. The clinical instance of REDCap (Clinical REDCap) was the third installment of REDCap at CAMH, preceded by a research instance and a fully validated instance for regulated clinical trials. REDCap version 9.1 was used for this implementation.

Technical Implementation

The technical implementation to integrate REDCap with Cerner included (1) tablet deployment in the clinic; (2) automated transfer of completed self-assessments in a PDF format to the patient chart; and (3) automated transfer of completed self-assessments in a discrete format to the patient chart for

XSL•FO
RenderX

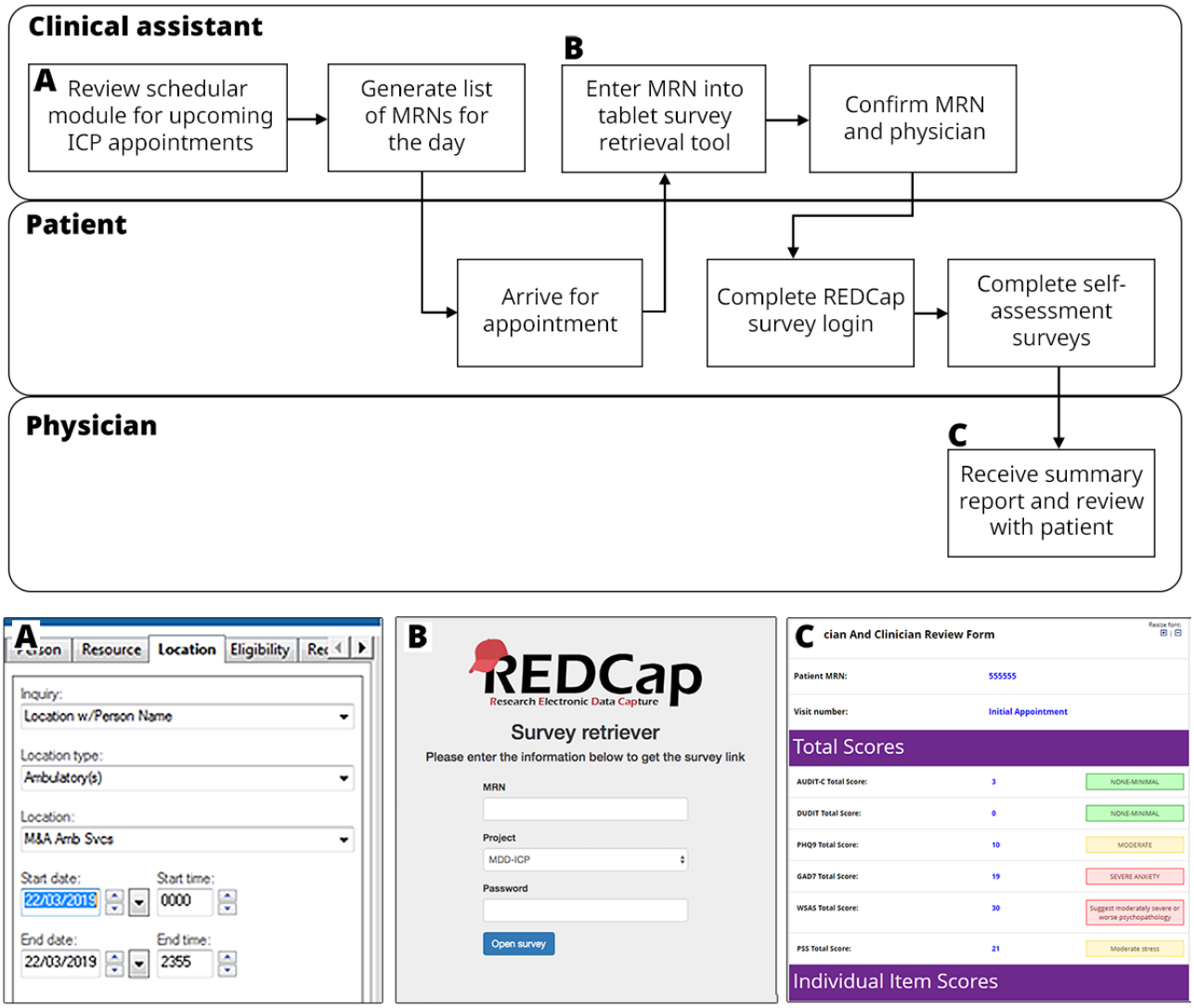
trending, analysis, and more complex physician decision support.

Tablet Deployment and Automated Survey Retrieval Tool

Samsung Galaxy Tab Android tablets were provided to clinical assistants to facilitate patient self-assessment data entry into REDCap. As a technical safeguard, a kiosk app (Kioware) was configured to limit tablet use to preapproved apps and webpages, in this case, Google Chrome and REDCap, respectively.

In collaboration with the clinic managers and physician teams, work was undertaken to streamline the administration of tablets and surveys for patients and staff. To simplify the clinical workflow (Figure 1), a custom *survey retriever* web app was developed using the REDCap application programming interface (API). This tool accepts the patient medical record number (MRN) and ICP name and retrieves the unique survey queue link from the *next available* set of self-assessments in the care pathway for a particular patient. Using this tool, the clinical assistant may load the appointment's set of assessments, beginning with a confirmation survey asking the clinical assistant to confirm the MRN, appointment date, and the patient's physician to ensure compliance and accuracy. The physician field on this confirmation form is populated using a dynamic SQL (Structured Query Language) field, which draws from a separate REDCap project that lists all ICP physician names or emails and their current status (eg, active or inactive). It allows clinical end users to easily update the physician list without needing to dynamically modify a dropdown list directly in the Online Designer. As a final patient verification step before data entry, the REDCap survey log-in feature is used to prompt the patient to enter a personal identification number (PIN) based on their date of birth.

Figure 1. Clinical REDCap workflow and example screens for a typical follow-up appointment. (A) Cerner scheduler view. (B) Custom survey retriever tool used as a tablet homepage. (C) Summary report built as a REDCap survey. ICP: integrated care pathways; MRN: medical record number; REDCap: Research Electronic Data Capture.



Transfer of PDFs to Cerner

To ensure the integrity of the health record, individually completed self-assessments were automatically transmitted as PDFs to the patient's EHR in Cerner and were filed in the *documents* section, specifically under *assessments and plans*. For convenience, an aggregate report is automatically generated within REDCap by piping values into a standalone REDCap

XSL•FO
RenderX

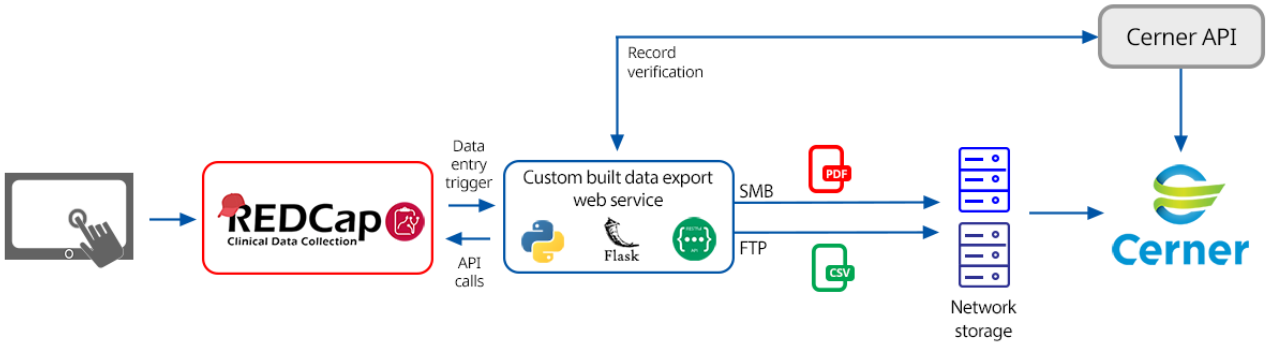
survey emailed to the primary clinician or physician upon assessment package completion (Figure 1). These reports summarize patient response data onto a single page; however, they are not sent to Cerner, as the data have already been transferred as individual files.

As the patient completes assessments, a REDCap data entry trigger (DET) delivers the project ID, record ID, event name, instrument, and form status to a custom-built data export web service. The service listens for DETs and sends a request back to REDCap, thus providing real-time data export functionality. Following the DET, the service makes a request to REDCap using the API method *export PDF file of data collection instruments* to retrieve a PDF document of the completed assessment (eg, the Patient Health Questionnaire-9 [PHQ-9]), and the document is written into a securely shared directory. Each PDF is given a unique file name that follows a predefined convention, as follows:

[mrn]_[encounter]_[assessment_code]_[date]_[time].pdf

This process ensures further compatibility with the Cerner interface, which uses a purpose-built Cerner AXRM COLD Feed (Cerner) [10] to import the document into the patient chart (Figure 2).

Figure 2. Clinical REDCap workflow and example screens for a typical follow-up appointment for a patient in the major depressive disorder integrated care pathway. API: application programming interface; FTP: file transfer protocol; REDCap: Research Electronic Data Capture; SMB: server message block.



Transfer of Discrete Data to Cerner

Similar to PDF export from REDCap, the transfer of discrete values also uses DETs and the REDCap API. In this case, the *export records* API method was used. This method returns the values of the completed assessment fields (eg, PHQ-9) and system variables, such as timestamps. Once the discrete data are retrieved, the web service executes the code to format this information as a comma-separated values (CSV) file that follows a specified convention, as follows: [timestamp].csv for the name and [mrn],[encounter],[uid],[assessment_code],[timestamp],[assessment_field1],[assessment_field2],...,[assessment_fieldN] for CSV content.

UID is a unique identifier of a completed assessment within the REDCap instance generated using the project ID, record ID, instrument code, and visit code. This identifier ensures that the correct form is updated in Cerner if the data in REDCap are modified. At this step, assessment variables are also renamed to be correctly mapped to the corresponding Cerner fields. The CSV file is then written to a directory on a file transfer protocol server. This directory is periodically scanned by a custom Cerner interface, which converts the CSV data into an HL7 unsolicited results (ORU) message. The interface then routes the HL7 to the standard Cerner external systems inbound servers for processing and passes the values into the results view of the chart, where the data are displayed in a table format. Currently, the interface only accepts numerical results, but future updates may include the processing of REDCap data sent in JavaScript Object Notation (JSON) to allow greater flexibility for data import.

Record Verification and Linkage

The primary verification mechanism for data transfer relies on correctly entering the MRN and a Cerner-specific appointment identifier called the *Encounter* number into REDCap. A custom Cerner API (Figure 2) was built, which accepts these parameters, checks whether these values are valid and belong to the same patient, and returns a JSON-formatted response. Upon submitting a data entry form in REDCap, a DET initializes a custom-built data export web service that sends the MRN and *Encounter* number to the custom Cerner API. If all the criteria are met and a successful response is received, the web service proceeds to create the document and discrete files. If the criteria are not met, the transfer is terminated, and an error message is logged and sent via email to the study contact. The message can be one of the following: “Missing MRN,” “Missing Encounter,” or “MRN and Encounter mismatch.”

Clinical REDCap Governance

To support technical development and integration, a robust and consistent governance model was necessary to ensure compliance with the legal health record and maintain the standardization necessary to effectively support scalable and reproducible MBC. We outline the governance practices adopted by our institution in a format that can be applied to other similar hospitals.

Clinical Governance Structures

The Clinical REDCap governance structure and principles were codeveloped with clinical and privacy stakeholders to establish core operating constraints. The underlying theme of these principles establishes Clinical REDCap as an extension of normal EHR data collection capabilities and establishes that

XSL•FO
RenderX

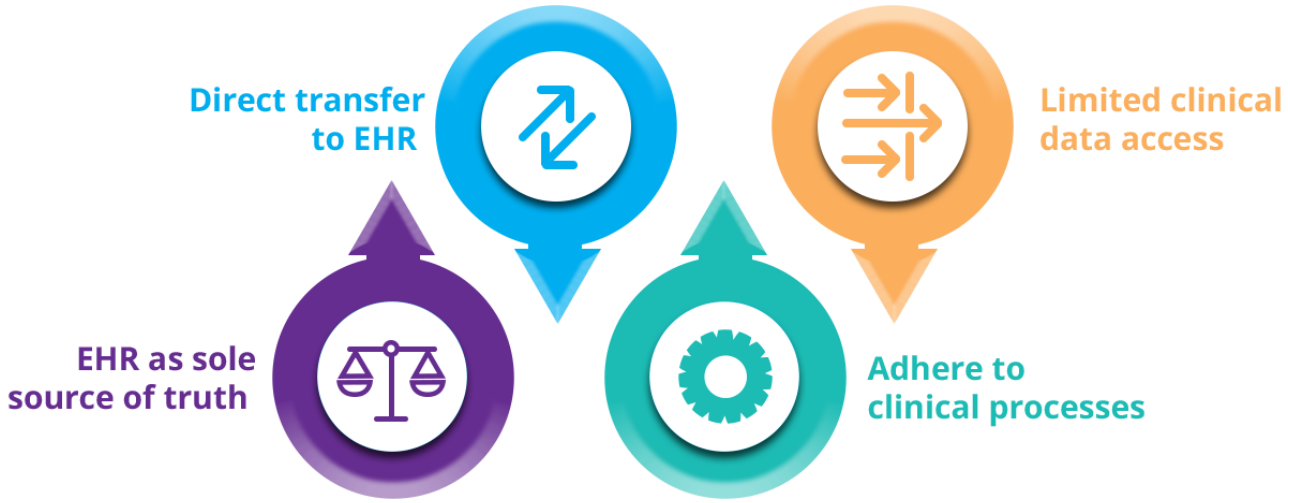
clinically relevant information must pass through the primary EHR to maintain a consistent health record. Changes to the care pathway within REDCap must be reviewed and approved by clinical governance committees, consistent with the review process for changes to Cerner. In general, institutional clinical governance committees are accountable for final decisions relating to platform governance and approval of forms used for clinical care; REDCap administrators are accountable for system administration and all project development, and the clinical applications (Cerner) team is accountable for final integration with the EHR. All parties were consulted or informed at various points where necessary.

The collaboration between all stakeholders established a list of 4 key operating principles (Figure 3) that were used to inform all decisions relating to Clinical REDCap, as follows:

1. The hospital EHR will remain the *sole source of truth*.
2. Clinical data collected in REDCap will be transferred directly to the EHR.
3. Clinical REDCap will follow existing clinical measure or pathway approval processes.
4. Access to data collected through Clinical REDCap will follow existing processes pertaining to clinical records.

Although the specifics of principles 3 and 4 will differ depending on the institution and regional legislation, the high-level principle of clinical governance alignment is strongly recommended.

Figure 3. Four key principles of REDCap (Research Electronic Data Capture) integration with the Centre for Addiction and Mental Health EHR. EHR: electronic health record.



Controlled Access and Data Retention

Access to the clinical instance is limited at both the server and individual roles or permissions levels. The clinical instance sever is situated behind the CAMH firewall and is not accessible externally. At the application level, user management is tightly regulated under structured clinical roles established by privacy and health record experts to meet the needs of the clinical workflows and limit unnecessary data access.

Users of the platform fall into 2 broad categories: administrator and clinical (Textbox 1). *Administrators* are responsible for project creation, instrument design, and API configuration and are strictly limited to central REDCap operations staff. In the *clinical* category, 3 subcategories exist: clinical data lead, a clinical staff member who acts as the *owner* of a project and provides in-clinic training and basic support (eg, data corrections); clinical care provider, the role assumed by clinicians and physicians who may enter new records and clinical data; and clinical assistant, a support staff who retrieve survey access codes, load survey queues, and supply tablets to patients for self-assessment. Account provision and project access are handled solely by administrators. Each new user must complete mandatory training by reviewing the policies of Clinical REDCap, including the differences between clinical and research platforms, permitted user roles, and privacy considerations when using the system. User roles are assigned following a hierarchical process: users who request a clinical data lead role must have their request approved by the institutional clinical director. For subsequent clinical roles, account or project access is approved by the respective clinical data lead, who also determines the user role (clinical care provider or clinical assistant) for the project. This approval is required before the user account or project access is provided by an administrator. This hierarchical approval process conforms to the internal organizational structure for consistency and varies in other institutions; however, this approach provides a template for developing permissions modes that are sufficiently restrictive but allow secure access as necessary.

At present, Clinical REDCap project data reside in REDCap for a period aligned with institutional clinical data retention periods unless otherwise removed by a REDCap administrator by request of the clinical team and privacy office.

XSL•FO
RenderX

Textbox 1. Clinical REDCap (Research Electronic Data Capture) user roles and responsibilities.

Clinical REDCap administrator

- Clinical REDCap staff responsible for account creation, modification, and suspension for all user types. Responsible for moving projects and draft changes to production. Limited to 2-3 staff members.

Project developer administrator

- Clinical REDCap staff responsible for creating REDCap projects and associated forms or surveys. Leads project validation to be carried out in collaboration with clinical stakeholders.

Service administrator

- Specialized account for managing secure services post data collection (eg, application programming interface data export to electronic health record).

Clinical data lead

- Clinical staff responsible for monitoring the integrity of the data, providing clinic-specific training to clinic users, supporting minor technical issues in the clinic, facilitating project validation, submitting project change requests, approving subordinate account requests, and serving as the main point of contact for developers and administrators.

Clinical care provider

- Clinical staff responsible for the administration and review of clinical assessments; can enter data that have been collected on paper into REDCap.

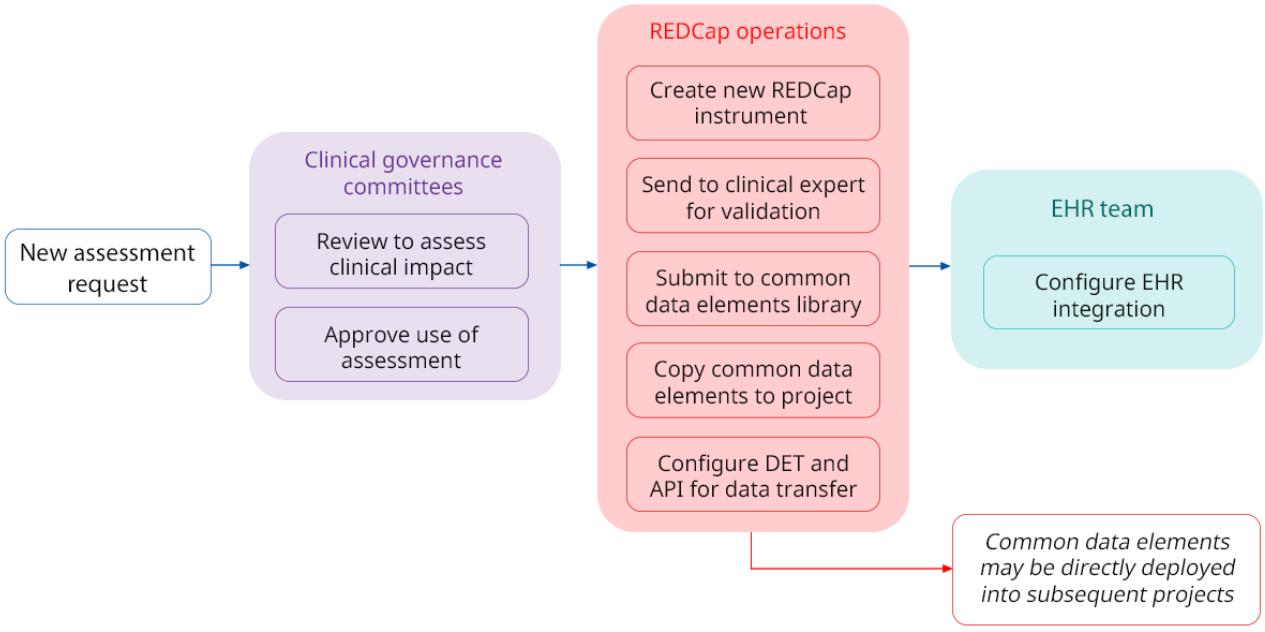
Clinical assistant

- Clinical support staff responsible for administering patient survey queue in the clinic and providing simple technical assistance.

Centralized Assessment and Project Development

All Clinical REDCap projects are built by a dedicated internal team of administrators that follow standardized naming and versioning practices for assessments and variables. All new assessments are clinically validated by CAMH clinicians or physicians and undergo strict approval processes through the hospital's clinical data governance committees to ascertain alignment with existing clinical instruments and standard practice. Once validated, assessments are deployed into a *common data elements* library to support the rapid development of parallel and complementary pathways (Figure 4). The use of common data elements across all projects ensures clinical consistency and provides a standard set of elements that can be transmitted to the hospital EHR with little additional setup. Ancillary forms such as project or participant tracking, reporting, or summary forms are built ad hoc as requested on a project-by-project basis.

Figure 4. High-level workflow for clinical instrument development on the Clinical REDCap instance. The review is conducted by clinical committees to align assessments. API: application programming interface; DET: data entry trigger; EHR: electronic health record; REDCap: Research Electronic Data Capture.



XSL•FO
RenderX

Evaluation of Impact

The clinical impact was assessed by determining the number of physicians who adopted the ICP before and after implementing the Clinical REDCap project. In addition, the monthly enrollment of patients was determined by counting the number of patients enrolled in the paper-based process compared with the Clinical REDCap process on average per month after 6 months.

A short user acceptance survey was developed and sent to Clinical REDCap users in the clinical care provider role to assess clinicians' and physicians' attitudes toward Clinical REDCap. The survey drew questions from the technology acceptance model, which seeks to measure users' perceived usefulness and perceived ease of use and how these factors may affect acceptance of new technology [11]. The survey consisted of 10 items measured on a 5-point Likert scale and included 3 free-text options for respondents to comment on their likes and dislikes and make suggestions for system improvement.

Results

Example Use Case: Major Depression Disorder Integrated Care Pathway

The major depression disorder integrated care pathway (MDD-ICP) provides a multidisciplinary and structured care management strategy, including interprofessional collaboration, standardized assessments, and evidence-based treatment algorithms. The pathway has been active at CAMH for 3 years, and a Clinical REDCap project was created as an alternative to paper-based data collection to digitize and support the clinic-wide implementation of the pathway in the fall of 2018. The data collected in Clinical REDCap (which are subsequently made available in the EHR) are used to support the MBC component of the MDD-ICP, which is used to inform treatment decisions.

Data Entry Workflow

Figure 1 shows a typical Clinical REDCap workflow for patients in the MDD-ICP. Every morning, a clinical assistant reviews the EHR scheduler module to compile a list of patients scheduled for an ICP appointment. When a patient arrives for their appointment, the clinical assistant retrieves a tablet and enters the MRN into the survey retriever tool. Provided that a valid MRN (ie, an existing record in REDCap) is entered, the retriever tool will open the first survey of the next available follow-up event. In this case, the first survey is a confirmation page that asks the clinical assistant to confirm the patient's MRN and their ICP physician(s). Next, the tablet is passed to the patient who enters the verification PIN (using the survey log-in feature in REDCap) based on their date of birth. Following successful verification, the patient may then complete their self-assessments, including the PHQ-9 [12], the Generalized Anxiety Disorder-7 [13], a substance use compliance screener, and a side-effects screener. Additional self-reports are completed at the initial and discharge appointments and include patient history, sociodemographics, the World Health Organization Disability Assessment Schedule 2 [14], the World Health Organization Quality of Life Instruments [15], and the Ontario Perception of Care [16] survey. In some use cases, the REDCap survey queue may be used to conditionally deploy certain assessments based on the responses provided in certain screener assessments. Once the patient has completed their assessments, a notification is emailed to the ICP physician with a link to a summary page in REDCap. This allows the physician to quickly review the results before and during the appointment without having to log into REDCap.

Impact of Digitization

The digitization of the MDD-ICP MBC component contributed to a 4-fold adoption rate by physicians in the clinic compared with the paper-based pilot implementation phase (Table 1). Digitization resulted in an efficient and streamlined process that automatically calculated assessment scores and the ability for psychiatrists to access and review results before a patient's appointment. This practice change, along with clinic-wide implementation, has contributed to the more than tripling of MDD-ICP patients seen by the clinic and has improved data quality. The use of Clinical REDCap ensures a standardized collection of assessments from patients at each visit and minimizes missing data. This systematic data collection method has resulted in the accumulation of high-quality data that may be used to inform quality improvement projects and to answer research questions with approval from the CAMH research ethics board.

XSL•FO
RenderX

Table 1. Overview of the impact and outcomes of Centre for Addiction and Mental Health major depressive disorder integrated care pathway digital transformation.

Measure	Digitization outcome	Impact
Physician adoption[a]	• Contributed to a 4-fold increase in-clinic physician adoption rate • Contributed to adoption by 20 physicians upon digitization compared with 5 physicians during the paper-based pilot implementation in the clinic	Increased capacity
Patient enrollment[a]	• Contributed to a 4-fold increase in total patient enrollment on average per month • After 6 months of digitization, an average of 49 patients enrolled monthly compared with 12 patients enrolled monthly during the paper-based pilot implementation	Increased capacity
Data entry time	• Elimination of paper transcription by support staff • Elimination of manual calculations	Increased workflow efficiency
Completeness	• Major reduction of missing values and incomplete assessment packages • Automatic date-stamping of all forms	Increased data quality
Standardization	• 3 new assessments submitted to the common data elements library for future deployment	Increased data quality

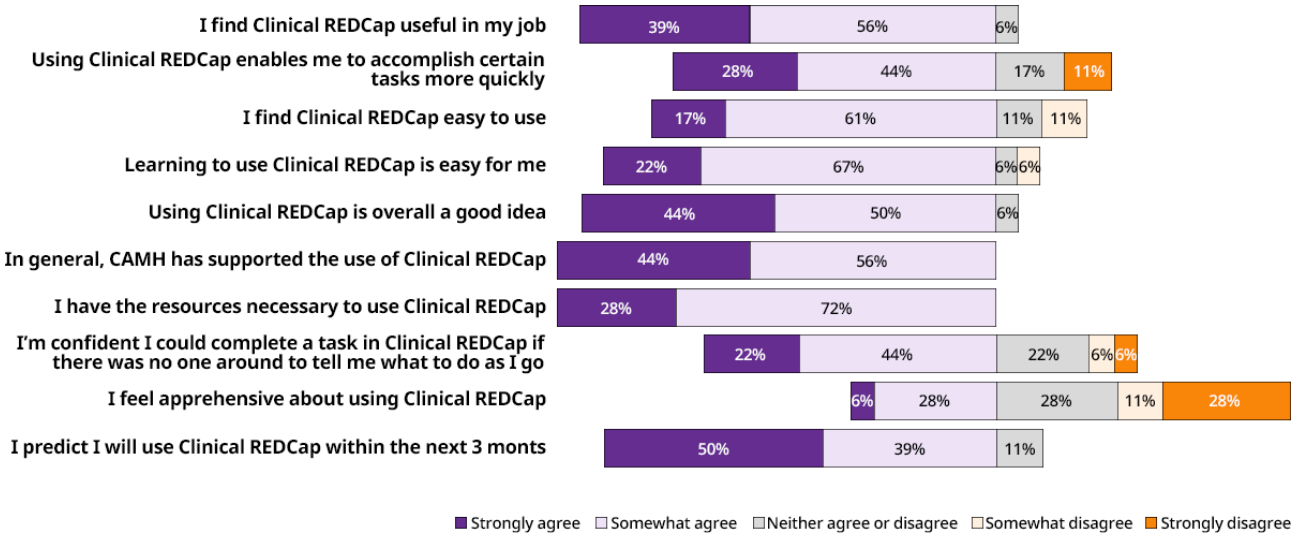
[a]Impact and outcome of clinic-wide major depressive disorder integrated care pathway implementation and digitization.

User Acceptance

All MDD-ICP clinical care providers were sent a 5-point Likert scale survey to assess their attitudes toward Clinical REDCap. A total of 18 responses were received, and Figure 5 shows the results of this survey.

Overall, respondents were satisfied with the system, with 94% (17/18) of respondents stating that they found Clinical REDCap useful in their job. Some apprehension was noted with ease of use, wherein 22% (4/18) to 33% (6/18) of respondents did not agree that the system was easy to use, that it allowed them to complete their work more quickly, or that they could complete a task without help. Despite these reservations, however, most respondents indicated that they planned to use the system within the next 3 months (use of the system is voluntary). From the free-text responses included in the survey, many respondents commented that electronic data capture presented major time savings compared with the previous paper-based entry method, especially in terms of the automatic calculation of scale scores. Conversely, many respondents noted that the interface could be more flexible and user-friendly with better tools for tracking patient trajectories.

Figure 5. Clinical REDCap user acceptance questions and responses. CAMH: Centre for Addiction and Mental Health; REDCap: Research Electronic Data Capture.



Discussion

Principal Findings

Ensuring sustained high-quality data collection is critical for supporting actionable MBC and clinical decision making. REDCap has proven to be highly valuable for deploying complex self-assessment workflows in diverse clinical environments. A governance model was established consistent with the HIMSS 7 designated processes that ensure regulatory compliance and integrity of the health record. System integration between REDCap and Cerner established real-time data flow for completed self-assessments.

Almost all clinical care provider users of Clinical REDCap indicated that they perceived the system to be useful; however, the survey revealed some concerns with the perceived ease of use of Clinical REDCap. These concerns are worth noting given

XSL•FO
RenderX

that a user's motivation to use technology is higher if the system is perceived as easy to use [17,18]. However, despite these concerns, all survey respondents indicated that they felt supported by the institution and that they had the resources available to use the system.

Efforts are underway to streamline workflows and improve integration and reporting. Currently, customized summary reports on questionnaire results are generated by Clinical REDCap and sent to physicians to review ahead of the patient's appointments for clinical decision making. This process has also had a positive impact on patient-physician communication and interaction during visits. However, the current presentation of self-report questionnaire results and EHR data focuses on point-of-care evaluation; the accessibility and presentation of complete treatment response trajectory over time would further enhance clinical decision support. To this end, enabling interactive data visualization methods (ie, dashboards) to convey knowledge about an individual's treatment trajectory would provide health care providers a tool for rapidly processing and interpreting the patient's chart and enhancing individualized clinical treatment decisions. Such a tool would free physicians from searching and filtering through the patient's chart, which can be challenging in a time-pressured clinical setting. Thus, both Clinical REDCap and EHR dashboards will be codeveloped with physicians using human-centered design methodology and embedded within the EHR to ensure streamline access during patient appointments. For the MDD-ICP, this dashboard will feature self-report measures trended over time along with corresponding treatment such as medication.

One of the core features of REDCap is the ability to send survey invitations via email, thereby allowing respondents to complete self-assessments at their convenience using their own devices. With the recent pandemic and temporary suspension of in-person clinic visits, the demand for such functionality (along with other virtual care solutions) has never been greater. An approach to this use case using Clinical REDCap may be configuring external access only to REDCap survey pages while prohibiting project backend access. This approach would allow off-site patients to complete their self-assessments while simultaneously protecting the project data and setup pages on the internal network. Although this strategy may precipitate novel privacy concerns surrounding the use of unsecured email (eg, sending to the wrong email address or guardians having access to their dependent's email account), this concern may be mitigated by using a patient verification PIN and other verification methods. Moreover, only blank self-assessment forms would be sent by email, and no sensitive health data or identifying information would be included. Consent language informing patients of these potential risks would be another key element in risk mitigation. A new process using this approach is currently being explored in several clinics that aim to send patients their self-assessment package 3 to 5 days before their appointment, followed by a reminder the day before. The self-assessment summary report can then be reviewed by the clinician or physician during a virtual appointment. It is anticipated that this new process will be able to leverage existing workflows and should not incur a significant increase in workload on the clinical staff compared with the original tablet-based approach.

The design of MBC pathways in Clinical REDCap will facilitate the rapid deployment of these pathways in other CAMH clinics. In addition, the scalable and extensible design of Clinical REDCap presents an opportunity for CAMH to partner with community sites to jointly implement MBC pathways in primary care settings. The use of assessments drawn from the common data elements library will enable standardized data collection practices consistent across the institution(s), thereby facilitating the accumulation of harmonized data that will allow clinicians and researchers to make meaningful comparisons over time and across otherwise independent clinical environments.

Conclusions

REDCap has been successfully deployed for the structured collection of standardized self-assessments to coordinate and accelerate the implementation of MBC in clinical practice. Systems integration was achieved between REDCap and Cerner to automatically transfer PDF and discrete variables from completed assessments into the patient's chart. This transfer maintained the integrity of the legal health record and enhanced clinician or physician decision-making ability. Thus, both a novel technical infrastructure and a comprehensive governance model aligned with institutional clinical committees were created to implement this new technology within the clinic. REDCap was demonstrated to be an effective tool for the implementation and expansion of MBC and ICPs by supporting an efficient and standardized process to deliver complex longitudinal bundles of assessments for the duration of a patient's treatment and by delivering real-time accessible information to health care providers to inform clinical decisions. This digitization process and integration has been designed to be extensible to other organizations practicing MBC in mental health and other disciplines.

Acknowledgments

This work was supported in part by the CAMH BrainHealth Databank, which is funded by the CAMH Discovery Fund. The work on the MDD-ICP was supported by the Max Bell Foundation and the University of Toronto Department of Psychiatry Academic Scholar Award. The authors would also like to thank the various clinical administrative staff members, clinical care providers, the clinical applications team, the enterprise data warehouse team, the information and privacy team, and the health records team for their work and feedback when designing and implementing this system.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019 Mar 01;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
2. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. Br Med J 1998 Jan 10;316(7125):133-137 [FREE Full text] [doi: 10.1136/bmj.316.7125.133] [Medline: 9462322]
3. Chen S, Awan S, Rajji T, Abdool P, George TP, Collins A, et al. Integrated care pathways for schizophrenia: a scoping review. Adm Policy Ment Health 2016 Sep 28;43(5):760-767. [doi: 10.1007/s10488-015-0696-z] [Medline: 26512011]
4. Settipani CA, Cleverley K, Hawke LD, Rice M, Henderson JL. Essential components of integrated care for youth with mental health and addiction needs: protocol for a scoping review. BMJ Open 2017 Apr 28;7(4):e015454 [FREE Full text] [doi: 10.1136/bmjopen-2016-015454] [Medline: 28455426]
5. Rotenberg DJ, Chang Q, Potapova N, Wang A, Hon M, Sanches M, et al. The CAMH neuroinformatics platform: a hospital-focused Brain-CODE implementation. Front Neuroinform 2018 Nov 6;12:77 [FREE Full text] [doi: 10.3389/fninf.2018.00077] [Medline: 30459587]
6. 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]
7. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
8. REDCap. URL: https://www.project-redcap.org/ [accessed 2019-08-21]
9. Sulkers H, Tajirian T, Paterson J, Mucuceanu D, MacArthur T, Strauss J, et al. Improving inpatient mental health medication safety through the process of obtaining HIMSS Stage 7: a case report. JAMIA Open 2019 Apr;2(1):35-39. [doi: 10.1093/jamiaopen/ooy044] [Medline: 31984343]
10. Shike H, Gaspari J, Newcomer L, Domen RE, Douventzidis T. P023 Direct transmission of HLA report to electronic medical record via cold feed interface. Hum Immunol 2016 Sep;77:57. [doi: 10.1016/j.humimm.2016.07.088]
11. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
12. 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]
13. 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]
14. Ustun TB, Kostanjesek N, Chatterji S, Rehm J, World Health Organization. Measuring health and disability : manual for WHO Disability Assessment Schedule (WHODAS 2.0). World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/43974 [accessed 2021-04-20]
15. Programme on mental health : WHOQOL user manual, 2012 revision. World Health Organization. 1998. URL: https://apps.who.int/iris/handle/10665/77932 [accessed 2021-04-20]
16. The Ontario Perception of Care Tool for Mental Health and Addictions (OPOC-MHA). Centre for Addiction and Mental Health. URL: http://improvingsystems.ca/projects/ontario-perception-of-care [accessed 2020-06-09]
17. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018 Dec 04;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
18. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011 Apr 28;9(1):46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]

Abbreviations

API: application programming interface
CAMH: Centre for Addiction and Mental Health
CSV: comma-separated values
DET: data entry trigger
EHR: electronic health record
HIMSS: Healthcare Information Management Systems Society
ICP: integrated care pathway
JSON: JavaScript Object Notation
MBC: measurement-based care
MDD-ICP: major depression disorder integrated care pathway
MRN: medical record number
PHQ-9: Patient Health Questionnaire-9

XSL•FO
RenderX

PIN: personal identification number
SQL: Structured Query Language

Edited by R Kukafka; submitted 10.11.20; peer-reviewed by E Portales-Casamar, L Rusu; comments to author 12.01.21; revised version received 03.02.21; accepted 16.03.21; published 20.05.21.

Please cite as:
Hawley S, Yu J, Bogetic N, Potapova N, Wakefield C, Thompson M, Kloiber S, Hill S, Jankowicz D, Rotenberg D
Digitization of Measurement-Based Care Pathways in Mental Health Through REDCap and Electronic Health Record Integration: Development and Usability Study
J Med Internet Res 2021;23(5):e25656
URL: https://www.jmir.org/2021/5/e25656
doi:10.2196/25656
PMID:34014169

©Steve Hawley, Joanna Yu, Nikola Bogetic, Natalia Potapova, Chris Wakefield, Mike Thompson, Stefan Kloiber, Sean Hill, Damian Jankowicz, David Rotenberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Agent-Oriented Goal Models in Developing Information Systems Supporting Physical Activity Among Adolescents: Literature Review and Expert Interviews

Kerli Mooses[1], PhD; Kuldar Taveter[1], PhD

Institute of Computer Science, University of Tartu, Tartu, Estonia

Corresponding Author:
Kerli Mooses, PhD
Institute of Computer Science
University of Tartu
Narva Maantee 18
Tartu, 51009
Estonia
Phone: 372 5526542
Email: kerli.mooses@ut.ee

Abstract

Background: Information and communication technologies (ICTs) are becoming increasingly popular in supporting the fight against low physical activity (PA) levels among adolescents. However, several ICT solutions lack evidence-based content. Therefore, there is a need to identify important features that have the potential to efficiently and consistently support the PA of adolescents using ICT solutions.

Objective: This study aims to create evidence-based models of requirements for ICT solutions supporting PA by combining scientific evidence from literature and health experts. In addition, we test the suitability of agent-oriented goal models in this type of modeling process.

Methods: A literature search of PubMed, Web of Science, and Scopus databases was conducted to identify evidence-based functional, quality, and emotional goals that have previously been proven to be relevant in supporting PAs among youth using ICT solutions. The identified goals were presented in the form of goal models. These models were used to collaborate with health experts to receive their input on the topic and suggestions for improvement. The initial goal models were improved based on the feedback from the experts.

Results: The results indicated that agent-oriented goal modeling is a suitable method for merging information from the literature and experts. One strength of agent-oriented goal models is that they present emotional requirements together with quality and functional requirements. Another strength is the possibility of presenting results from a literature review in a systematic manner and using them thereafter in the communication process with stakeholders. Agent-oriented goal models that were created were easy to understand for health experts without previous experience in requirements engineering, which facilitates and supports collaboration with nontechnical stakeholders.

Conclusions: The proposed agent-oriented goal models effectively merged information from scientific literature and experts in the field and presented early functional, quality, and emotional requirements in a holistic and coherent manner. We believe that the created models have high potential to help requirements engineers and developers to provide more efficient ICT solutions that support PA among adolescents in the future.

(J Med Internet Res 2021;23(5):e24810) doi:10.2196/24810

KEYWORDS

agent-oriented goal models; physical activity; adolescent

XSL•FO
RenderX

Introduction

Background

Despite the numerous health benefits of physical activity (PA) on mental, social, and physical health [1-3], countries struggle with low levels of PA among children and adolescents [4]. According to data from 32 European and North American countries, only 23% of adolescent boys and 14% of girls meet the recommended daily PA levels [4], which is a minimum of 60 minutes of moderate- to vigorous-intensity PA every day [5]. According to the studies, PA declines during adolescence [6,7]. At the same time, many health-related behavior habits are established during adolescence that carry on to adulthood [8,9]. Therefore, adolescence is a sensitive period that can influence health behavior in later life, and greater attention should be paid to supporting the PA of adolescents [9].

Numerous interventions have been developed to support PA among children and adolescents, with most of them being conducted in a school setting [10], where it is easy to reach all children despite their age or socioeconomic background. In recent years, interventions using a variety of technological solutions have emerged, which can be explained by the high level of technology use among children and adolescents as well as by the vast possibilities of information and communication technologies (ICTs) that enable more individualized behavior change interventions [10]. It is also important that through ICT solutions, it is possible to provide health information to people who would otherwise not have access to health education resources [11]. According to adolescents themselves, supporting PA by using technology makes PA more appealing and helps to increase their awareness of the actual activity levels [12]. Previous ICT interventions have shown promising results [10,13-20], irrespective of the technological mechanisms used to deliver or support the delivery of an intervention. Some examples of the mechanisms used are the internet [14,15,21], text messaging, email [19-21], and phone apps [22]. However, the development process of ICT solutions supporting PA includes several challenges, which is highlighted by the fact that not all ICT interventions have been proven to be effective or successful [22,23]. It has been pointed out that a stronger collaboration between scientists and technologists who develop the solutions is needed [22,24,25], as ICT interventions are more likely to be successful when combined with theory, such as behavior change theory [15,21]. Collaboration with health experts and scientists during the problem identification and requirements engineering phase helps to elicit evidence-based functional, quality, and emotional requirements for ICT-based intervention solutions that have the potential to support the behavioral change of the target group. Such collaboration with scientists and experts would also increase the quality of information, and according to a review, several apps lack evidence-based content and focus mainly on functionality, aesthetics, and engagement [22].

In addition to functional requirements, the quality and emotional side of the software application must also be considered because the adaptation of the application depends on the emotions of the user [26]. The lack of user-friendliness and being unappealing to the target group has been marked as a significant cause of dropout and low levels of involvement in previous interventions [23,27]. Therefore, it is crucial that the development process includes both end users [27] as well as different stakeholders, health experts, and scientists, which ensures that the end product will have an expected impact on the end users [25,28-30]. The method used to involve different stakeholders must be easy to comprehend, as stakeholders often lack previous experience in requirements engineering. At the same time, it should capture as many functional, quality, and emotional requirements that constitute an important input for requirements engineers and developers.

Agent-Oriented Goal Modeling

One method for presenting emotional requirements for sociotechnical systems (STSs) is agent-oriented goal modeling [29-32]. STS is a software-intensive system that has defined operational processes followed by human operators and operates within an organization and comprises both social and technical aspects [33]. The STS consists of humans, software, and hardware [34]. According to the preceding definitions, ICT solutions supporting PA among adolescents fit into the category of STSs. Agent-oriented goal modeling [31,32,35] is a state-of-the-art method for eliciting and representing the requirements of STSs. In this method, the starting point is something to be achieved or done, whereas it is not yet important *who* or *what* does it. A distinct feature of motivational modeling is its explicit support for eliciting and representing emotional requirements [30,36]. The method enables the representation of functional, quality (nonfunctional), and emotional requirements for STSs in a holistic manner as a hierarchical goal tree, consisting of functional, quality, and emotional goals [30-32,35,36]. Agent-oriented goal modeling has been applied previously in several studies, which have shown that this approach is suitable and comfortable for including both technical and nontechnical stakeholders [30,31,36-38]. Moreover, this approach supports communication with nontechnical experts and other stakeholders whose input is required for the design process. Agent-oriented goal models provide a thorough description of possible solutions for the problem in question. The created models are not final technical requirements but rather offer an overall picture of the necessary elements and features that should be considered during the development process. On the basis of the agent-oriented goal models, requirements engineers can form more concrete requirements in the form of, for example, user stories [39], which can then be used in the further development process. Often, the creation of goal models includes a workshop or focus group with stakeholders to identify functional, quality, and emotional goals and the roles required for the attainment of the goals [30-32,36], which is followed by creating goal models based on the information gathered. However, in the current situation caused by the spread of COVID-19, the possibilities for face-to-face workshops have become limited, and alternative solutions have to be considered. In addition to expert opinions, more scientific evidence is emerging with regard to the features and factors that increase the effectiveness of ICT solutions supporting PA, which should also be considered in goal models.

XSL•FO
RenderX

Objective

The main aim of our study is to create evidence-based models for ICT solutions supporting PA by combining scientific evidence from the literature and opinions of health experts. For this purpose, we also aim to pilot the use of agent-oriented goal modeling in the process of communicating the results of the literature review and involving experts with nontechnical backgrounds. Interestingly, we have previously used a similar approach in problem domains where arranging co-design workshops would be complicated, impossible, or even dangerous [37,38].

Methods

Overview

Previous studies using agent-oriented goal models have often gathered their input for goal models only from stakeholders through interviews or workshops [30,32,35,36]. For the topic under discussion, there is some scientific evidence on the features that have the potential to increase the effectiveness of ICT solutions supporting PA among adolescents and reduce dropout. Therefore, to combine the findings from previous ICT intervention studies with information from experts, this study was designed in the following steps:

- Identification of evidence-based features of ICT solutions that have the potential to support the behavioral change and PA of adolescents based on a literature review
- Development of initial goal models
- Inclusion of experts, model validation, and improvement

Literature Review

First, we identified from the scientific literature evidence-based features of ICT solutions that have the potential to support the behavioral change and PA of adolescents. For this purpose, we searched for articles in the PubMed, Web of Science, and Scopus databases. Separate searches for three topics were conducted for each database. The topics and keywords used are listed in Textbox 1. The first search focused on reviews that covered ICT-based interventions with adolescents. The results of the first search indicated that behavior change techniques (BCTs) have the potential to positively influence the behavior of the target audience [15,21]. Considering this, the following search focused on identifying the most prevalent and effective BCTs used in ICT solutions aimed at adolescents. To gain more thorough insight into the needs and preferences of adolescents concerning ICT solutions and PA, the third search focused on qualitative studies. The reference list of each identified article was also searched for relevant reviews and studies.

The reviews that addressed ICT-based interventions with adolescents included different study methodologies, such as randomized controlled trials, quasi-experimental studies, and feasibility studies [14,15,17-21,24,40-43], providing a good overview of the features that have the potential to support PA levels. Information about BCTs was mostly elicited from studies or reviews describing ICT solutions using BCTs [22,25,44-46]. Finally, qualitative studies helped to identify possible facilitators and barriers perceived by the adolescents themselves [7,12,23,47-58].

Textbox 1. Search topics and keywords.

Search: reviews covering information and communication technology–based interventions with adolescents

- "(adolesc*[Title/Abstract]) AND (review[Title/Abstract]) AND ('physical activity'[Title/Abstract]) AND ((technology[Title/Abstract]) OR (web[Title/Abstract]) OR (E-Health[Title/Abstract]) OR (eHealth[Title/Abstract]) OR (mHealth[Title/Abstract]) OR (app*[Title/Abstract]))"

Search: behavior change techniques used in information and communication technology solutions aimed at adolescents

- "((adolesc*[Title/Abstract]) OR (young [Title/Abstract])) AND ('physical activity'[Title/Abstract]) AND ('behaviour change'[Title/Abstract])"

Search: qualitative studies to identify factors valued by adolescent users

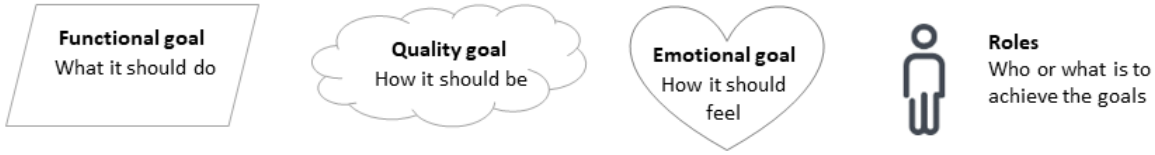
- "('physical activity'[Title/Abstract]) AND ((qualitative[Title/Abstract]) OR ('focus group'[Title/Abstract])) AND ((app*[Title/Abstract]) OR (mHealth[Title/Abstract]) OR (eHealth[Title/Abstract])) AND (adolesc*[Title/Abstract])"

Development of Initial Goal Models

The data gathered from the literature were used to create the initial version of goal models, which included three types of goals: *do*, *be*, and *feel* goals. *Do* goal is a functional goal that indicates what the system should do. *Be* goals or quality goals present the nonfunctional requirements of the system, which describe the quality aspects of the system—how the system should be. *Feel* goals are emotional goals that describe how the user should feel when interacting with the system. In addition, *roles* are identified, which represent who is responsible for the attainment of which goals. This information is eventually presented as a simple graph in which the functional goals are rendered in a tree-like hierarchy. In the hierarchical goal model, each subgoal represents a particular aspect of achieving its parent goal. The functional goals are represented with tilted rectangles, whereas roles, quality, and emotional goals are attached to the appropriate functional goals and are represented by stick man icons, clouds, and heart symbols, respectively (Figure 1). Therefore, it is important that roles, quality, and emotional goals attached to the given functional goal apply to that goal and to all of the functional goals located below it in the goal tree. The described method enables the representation of functional, quality (nonfunctional), and emotional requirements for an STS in a holistic manner [30-32,35,36].

XSL•FO
RenderX

Figure 1. Notation for goal modeling.



The initial functional, quality, and emotional goals identified in the literature are presented in Textbox 2.

The features identified from the literature were combined into a hierarchical goal model, with *move* as the highest-level goal—purpose of the ICT solution—and *report background info*, *set goals*, *monitor activity*, *receive feedback*, and *increase awareness* as first-level functional goals. These first-level functional goals were elaborated into lower-level functional goals and attached to their quality and emotional goals. The initial goal model derived from the literature review was presented to the experts, who elicited additional functional, quality, and emotional goals. In the final goal models, the functional, quality, and emotional goals identified from the literature are marked with an asterisk (*), and those elicited by the experts are marked with a number sign (#).

The goal models can be further elaborated into the format of user stories [39], which is one of the most prevalent ways of representing requirements in agile software engineering [73,74]. We made use of the following format of user stories, which has been adapted from Cohn [73]: “As a [user performing a certain role], I need [to perform action] to support [achieving a certain goal]”. We will introduce an example of elaborating goal models into user stories in the *Elaboration to User Stories* subsection.

XSL•FO
RenderX

Textbox 2. Initial goals identified from the literature.

Functional goals

1. Report background information [20,40,44,59-61]
2. Set goals [21,24,42,45,52-54,60,62]
3. Monitor activity [21,24,42,45,62]
4. Receive feedback [21,24,42,45,51,54,62]
5. Participate in challenges [44,46,53,63]
6. Increase awareness [54,60,64,65]
7. Select challenge [12,47,50,66]
8. Create groups [15,48,63,67-69]
9. Invite friends [15,48,64,67-69]
10. View results [55,60]
11. View comparison with others [46,55,56,60,64]
12. Show health benefits [22,55]
13. Receive activity suggestions [22,55]
14. Receive links to materials [22,55]
15. Help to interpret the data [52,64]

Quality goals

1. Actively [5]
2. Regularly [5]
3. Personal [21,45,46,51,52,60]
4. Achievable [12,47,50,53,66]
5. Challenging [12,47,50,66]
6. Scientific [18,20,22,54,60]
7. Supportive [22]
8. Attractive [43,46,54,60,69]
9. Easy-to-use [43,46,54,60,69]
10. Automatic [43,46,54,60,69]

Emotional goals

1. Enjoyment [47,49,50,53,70-72]
2. Fun [47,49,50,66,70-72]
3. Involved [44,46]
4. Feel achievement [53,66]
5. Engaged [44,46,53,56,64]
6. Thrilled [53]
7. Be aware [18,20,22,54,60,64,65]

Inclusion of Experts, Model Validation, and Improvement

The initial goal models were validated by the experts. Although previous studies with goal models have used workshops or focus groups to gather input from stakeholders [30,32,36], due to the COVID-19 situation, in this study, one-on-one interviews in the outdoors and web-based meetings were used instead. We consulted sports scientists and physical education experts (n=5) to identify the features of ICT-based interventions that can support the PA of adolescents. Some of the included experts had previous experience in the development and implementation of PA interventions in a school setting; therefore, they could highlight several aspects and experiences from the field. All the included experts were also parents, which enabled them to reflect the expectations of parents. The initial goal models were

XSL•FO
RenderX

sent to the experts before the interviews, together with an explanation of the notations and the aim of the models. These aspects were repeated by the interviewer at the beginning of each interview. Thereafter, the interviewees' thoughts about the goal models rooted in the literature and their possible improvements were obtained. All suggestions provided by the experts were documented by the interviewer in written form. In addition, feedback concerning the clarity, usability, and comprehensiveness of the agent-oriented goal models was provided. All interviews were conducted by the same researcher.

The final versions of the goal models resulting from both the literature review and the interviews with the experts are presented and explained in the *Results* section. Goals in goal models include markings that distinguish between the ideas originating in the literature and the ideas proposed by the experts. As both the ideas appearing in the literature and the ideas proposed by the experts reached a detailed level, the resulting goal models also cover the functionalities of ICT solutions that are usually represented as detailed requirements in the form of user stories [39]. In the goal models presented in the *Results* section, such functionalities are distinguished with a different color of rectangles denoting functional goals. For completeness, in the *Participate in Challenges* subsection, we also present user stories.

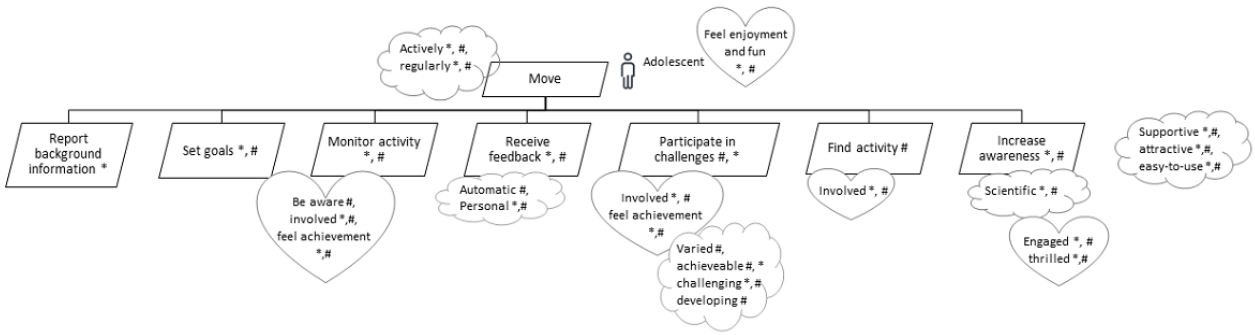
Results

Overview

The outcomes of the literature review, complemented according to the opinions expressed by the experts, were represented by the goal model shown in Figure 2. The highest-level goal expressed by the goal model is that adolescents should move or, in other words, be engaged in PA. In terms of health benefits, PA should be performed actively and regularly because according to the recommendations for PA [5], adolescents should acquire every day a minimum of 60 minutes of moderate-to vigorous-intensity PA. This is reflected by the *actively* and *regularly* quality goals associated with the main goal. We also attached the emotional goals *feeling enjoyment* and *feeling fun* to the main goal, as there is an abundance of evidence that enjoyment and fun contribute to increased PA among children and youth [47,49,50,70-72]. The highest-level goal was elaborated into seven subgoals, each representing a particular aspect of achieving the highest-level goal, together with the associated quality and emotional goals.

In the following five subsections, we will explain the seven second-level functional goals represented in Figure 2 and their subgoals along with the quality and emotional goals pertaining to the functional goals.

Figure 2. Overall goal model for information and communication technology solutions supporting physical activity levels of adolescents. The "*" in the model represents findings from the literature; the "#" represents suggestions by experts.



Report Background Information

The first activity when starting to use an ICT solution should be providing background information to offer personalized experience based on the adolescent's personal objectives, previous PA experience, age, or gender. Using information provided by the adolescent allows the provision of messages that are more relevant to the recipient, thereby reducing information overload and supporting long-term engagement [44,59-61], while receiving personalized encouraging messages increases the effectiveness of the intervention [20,40]. A qualitative study with students confirmed that tailored information is preferred by young users [60]. Having sufficient background information can be applied for nudging [75] and directing users to healthier choices.

Set Goals, Monitor Activity, and Receive Feedback

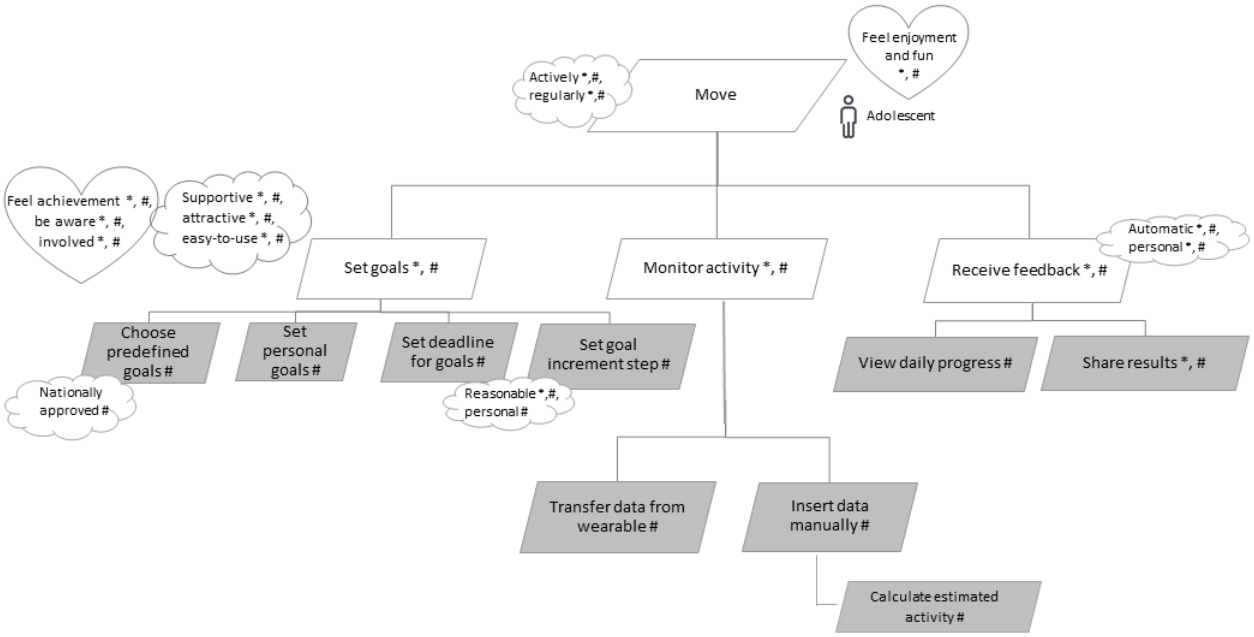
The inclusion of BCTs in ICT solutions for supporting PA has been emphasized by several authors, as they seem to have a significant impact on the effectiveness of the solution [15,21,22,40,41,45]. However, there is no consensus on the optimal number of BCTs that should be included in an effective ICT solution. It has been pointed out that the inclusion of multiple BCTs increases the attractiveness and effectiveness of the app [22,76]. At the same time, some reviews with adults indicate that too many BCTs can have a negative impact on effectiveness [62,77]. On average, six BCTs have been included in the apps aimed at children and adolescents [22], with the most popular BCTs being *provide instructions*, *provide general encouragement*, *provide contingent rewards* [22], *prompt specific goal setting*, *prompt self-monitoring of behavior*, and *provide feedback on performance* [21,45]. The importance of setting goals, monitoring activity, and receiving feedback was also highlighted by the experts involved in the modeling. Adding these three BCTs to the ICT-based solution can have a positive impact on the effectiveness of the solution [24,42,62]. The experts also pointed out that the inclusion of the functionalities *set goals*, *monitor activity*, and *provide feedback on performance*

XSL•FO
RenderX

makes it possible to incorporate the ICT solution into teaching activities at school (eg, physical education and health education lessons), which would have multiple advantages. For example, in addition to helping students gain knowledge about their actual activity levels, incorporating the solution into in-school or in-curricula activities has the potential to increase its effectiveness [14,21] and ensure that less-motivated students also use the solution [15].

On the basis of the literature review and feedback from the experts, we added *set goals*, *monitor activity*, and *receive feedback* as the second-level functional goals. The emotional goals associated with these functional goals are *be aware*, *feel achievement*, and *involved*. The second-level functional goals are elaborated into sublevel functional goals, as shown in Figure 3.

Figure 3. Sublevel models for the functional goals "set goals," "monitor activity," and "receive feedback." The "*" in the model represents findings from the literature; the "#" represents suggestions by experts; gray rectangles represent functionalities at the level of user stories.



For the second-level functional goal, *set goals*, shown in Figure 3, we identified that it should be possible to select from among predefined goals that are based on national and international PA recommendations as well as define one's own goals. The benefit of defining personalized goals is that such goals match the user's abilities and daily routines [46,51,52,60] and offer a sense of enjoyment from achieving one's own goals [53]. Personalized goals are especially useful when goals based on national or international PA recommendations seem to be too demanding and, because of this, constantly failing to reach them would be demotivating. It is important that the goals that have been set are achievable because perceived competencies are associated with the feeling of fun [66], which, in turn, supports the continuous usage of the ICT solution and behavior change. Therefore, defining one's goals that seem realistic and achievable supports the self-efficacy, autonomy, and motivation of adolescents. The importance of providing goal setting functionality was also highlighted by students and adolescents [52-54,60]. The experts pointed out that the possibility of setting goals enables the solution to be incorporated more easily into a school setting and use it in physical education lessons. As achievement goals should be accompanied by deadlines and sometimes it is beneficial to set subgoals with reasonable increment steps, we believe that such functionalities should be included in ICT solutions.

Goal setting functionality is strongly associated with the self-monitoring functionality to evaluate compliance with the goals that have been set. In Figure 3, this functionality is represented by the *monitor activity's* second-level functional goal. On the basis of the literature, self-monitoring seems to be a functionality that is crucial in ICT solutions aimed at supporting PA levels [24,42,54,60,62]. Young users value self-monitoring as it helps to increase their awareness of the actual PA levels and review their progress over time [54,60,64,65], which, in turn, has been associated with increased PA levels [43,64,78]. There are many ways to measure one's PA by, for example, using subjective ratings and questionnaires or objective data from pedometers, accelerometers, or heart rate monitors. The strength of transferring data from wearable trackers is to provide objective PA information that is more accurate compared with the information obtained by means of questionnaires and more comfortable gathering. The ICT solution should connect with wearables from different manufacturers and convert the data into an easily understandable format. However, activity trackers might underestimate some activities (eg, skiing and cycling), or it can be impossible to wear them during activities such as swimming and wrestling. In addition, users sometimes forget to wear the tracker or forget to start tracking at the beginning of the workout [55]. Therefore, it is possible to manually enter training or activity data into the ICT solution.

The third second-level functional goal shown in Figure 3—*receive feedback*—is tightly associated with goal setting and monitoring PA to evaluate performed activities and plan

XSL•FO
RenderX

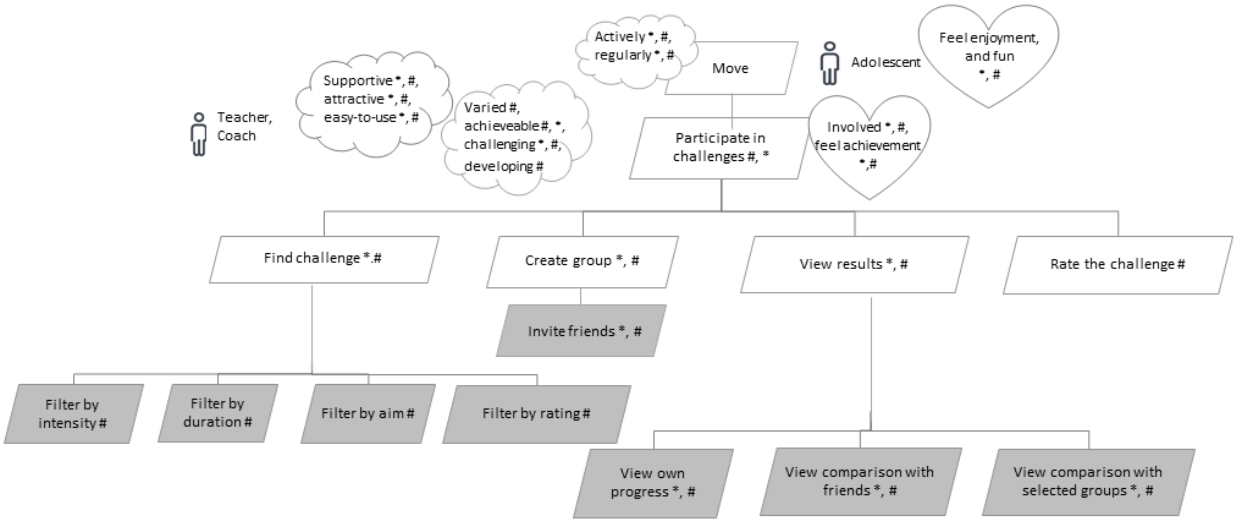
future ones. Providing feedback is a necessary feature pointed out by young app users [51,54], and one way to increase engagement with the solution [44] is to enhance the user's motivation and self-efficacy [46]. Knowing that healthy PA levels have been reached increase the feeling of being healthy among adolescents [52]. However, the way feedback is given should be carefully planned and targeted so that it would not bring about negative emotions, such as guilt and disappointment [60], which can discourage and cause dropout.

Sharing results was also added to the model as one of the functional goals because sharing can increase the motivation of adolescents by receiving social support. Sharing also increases the feelings of belonging [46,56,64] and contributes positively to the effectiveness of the solution [40]. According to the experts involved in the modeling process, sharing can widen the usage of an ICT solution because it can be incorporated into physical education lessons by sharing the activity goals and results with teachers. In such a case, the physical education teacher can also serve as the provider of tailored feedback, thereby increasing the motivation of the students to be physically active. However, this possibility should be prevalidated with adolescents, as it might be perceived by them as a controlling mechanism resulting in a loss of autonomy, which can, in turn, cause resistance to using the ICT solution by adolescents [52].

Participate in Challenges

Challenges and competitions have been found to be the elements of game design that can increase engagement and reduce dropout [44,53]. In addition, participating in challenges and competitions increased the PA of the participants [46,63]. At the same time, not all people are interested in participating in challenges and competitions [12,40,46,52,53]. The studies indicate that participation in challenges is associated with PA and self-efficacy levels, as those with higher activity levels are more prone to participate in challenges [46]. It has been pointed out that competitors should have similar precompetition activity levels so that the competition effect would not drop [63] as a constant feeling of failure and no hope of winning reduces the motivation of participants [52,53]. Therefore, in addition to finding challenges, it should be possible to create groups of participants with similar PA levels or skills. This is reflected in the goal model shown in Figure 4. In addition, the ICT solution should include a variety of challenges for adolescents with different PA levels, interests, and skills, which make the participants feel competent [12,47,50,66] and would support their feeling of autonomy [50]. Moreover, the challenges have to be versatile, as similar challenges become boring for adolescents [53].

Figure 4. Sublevel models for the functional goal "participate in challenges." The "*" in the model represents findings from the literature; the "#" represents suggestions by experts; gray rectangles represent functionalities at the level of user stories.



As challenges that extensively prioritize only successful performance can create subgroups of nonparticipants [56], it was suggested by the experts that the challenges should involve different kinds of targets. For example, in addition to finding the most active participants, an achievable daily threshold of a goal set by the adolescent should also be established with the aim of reaching it on all or most of the days comprised by the challenge. This approach would provide the participants with the possibility to feel a sense of achievement, increase self-efficacy, and retain anticipation to reach the goal. This also complies with the views expressed by the adolescents that not only excelling in particular sports but also merit of effort should be awarded or noticed [12]. Moreover, focusing on achieving healthy daily PA levels instead of maximum performance helps to avoid extreme PA behaviors that have been reported by some adolescents as a result of participating in PA challenges [52].

Participating in challenges is a social activity that involves interactions with other participants. The evidence suggests that participation in the challenges might be more effective when friends are engaged because there is an abundance of evidence that both the motivation to be active and actual PA levels are increased when friends are involved [15,48,67-69]. Moreover, it is more fun to be physically active together with friends [47], and peer involvement has been previously identified as a feature

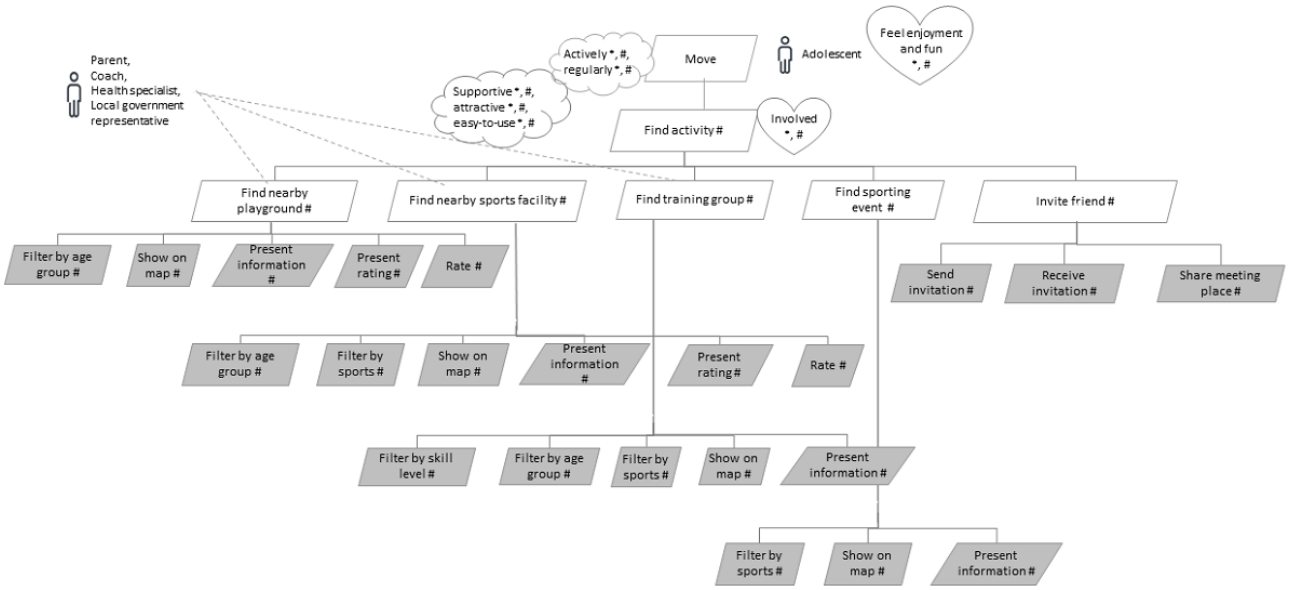
XSL•FO
RenderX

that adolescents like and use in ICT solutions [64]. Therefore, one should be able to invite a group of friends with whom to participate in the challenge. This is reflected by the corresponding fourth-level functional goal, *Invite friends*, shown in Figure 4. Another way of interaction during the challenge is to share information with coparticipants or a selected group of people. This is represented in Figure 4 by the third-level functional goals *view results* and *rate the challenge*. On the basis of the literature review, sharing PA information with familiar people supports PA levels, whereas sharing with strangers seems to be ineffective [46]. There are mixed feelings about comparing the results of meeting the challenge, similar to the case of participating in challenges. Some users of ICT solutions have found sharing demotivating when a comparison is made with more active people, and they rather prefer to compare themselves with lower performers to obtain assurance and confidence [55,60], whereas others find the comparison with more active people motivating [55]. According to the suggestions of the experts, we also complemented the goal model shown in Figure 4 with the roles *teacher* and *coach*, so that performers of these roles could incorporate suitable challenges into their teaching activities and would be able to view the outcomes.

Find Activity

According to a socioecological model [79], one's PA is also influenced by the built environment. To increase the awareness of PA opportunities available in the neighborhood, we have added, based on the literature, the functional goals *find nearby playground*, *find nearby sports facility*, *find training group*, and *find sporting event*. These functional goals along with their subgoals are presented in Figure 5. Although university students considered the feature of providing information about facilities and possibilities provided by the environment as unimportant [55], the experts involved in our study suggested that compiling a comprehensive overview of the PA opportunities in the neighborhood can support PA because free and easily accessible sports and activities are more appealing to adolescents [12]. By means of this feature, adolescents can also find places to practice nonorganized and lifestyle sports activities, such as workouts, slackline, parkour, or skateboarding, which are becoming increasingly popular among youth [80]. At the same time, this type of feature can be beneficial for different stakeholders. For example, among those who can benefit from such features are parents who want to be active with their children, teachers and coaches who want to find training places, coaches who want to be visible to their target group, and experts in public health who would like to introduce different PAs to inactive adolescents. In addition, representatives of the local government can also benefit from these features, as the ICT solution can help to identify the existing opportunities for PA and plan additional developments, considering that both proximity and accessibility of facilities affect participation in PA [47,57,66]. In addition, the presence of green spaces has the potential to increase the PA of children [81]. Therefore, the *find activity* feature would help the local government to apply nudging [75] by making healthy choices more accessible to the target group. Providing a neighborhood with opportunities for PA also reduces children's dependence on their parents, which has been identified as one of the factors hindering adolescents from participating in sports [12].

Figure 5. Sublevel models for functional goal "find activity." The "*" in the model represents findings from the literature; the "#" represents suggestions by experts; gray rectangles represent functionalities at the level of user stories.



It should be kept in mind that the preferences of adolescents vary greatly—some prefer competitive and some others prefer noncompetitive activities, and some prefer team sports and some others prefer individual sports [12,47,58]. Moreover, nonorganized and lifestyle sports activities are gaining popularity among youth [80]. To support the PA of all adolescents, different types of sports with different levels of competitiveness should be offered, as providing a choice can increase the motivation and willingness to participate [50,56]. According to the qualitative studies, adolescents value activities that are diverse, challenging, social, and support their autonomy [50] and offer the possibility to try new activities [12,82].

XSL•FO
RenderX

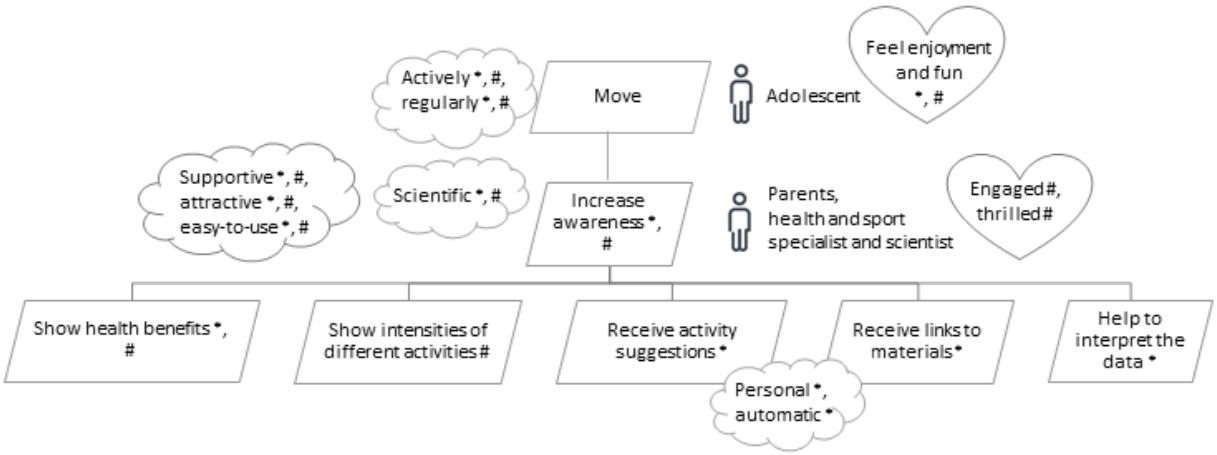
Adolescents should be able to distinguish between training groups based on their expected level of skills (eg, beginners, advanced, or high level) because joining a training group with more skilled participants lowers self-efficacy, reduces motivation, and increases the odds of dropout.

Increase Awareness

Current apps aimed at youth lack health content and concrete recommendations [22,25], and only a few apps are aligned with PA recommendations [22]. The need for evidence-based content has been emphasized both by researchers [18,20,22] as well as by users of ICT solutions [54,60]. Therefore, one focus of an ICT solution should be presenting the benefits of PA. This need is reflected by the functional goal *increase awareness* and its subgoals, along with the relevant quality and emotional goals, as shown in Figure 6. This should not only include the benefits for physical health but also the benefits for social and mental health and academic achievements [1], which are less known to the wider public according to the experts who were involved in the goal modeling process of our study. In addition, developing positive attitudes toward PA and informing stakeholders on possible health gains can increase adherence to ICT solutions [24,61]. Students also reported that they would like to receive tips on how to reach their achievement goals, make activities more fun, exercise safely, and get recommendations on when it is best to exercise [55].

The feature of increasing awareness should also include providing advice on the interpretation of the collected data, as adolescents tended to neglect the information about their PA behaviors that they did not understand [52]. Providing support for data interpretation, especially at the outset of using the ICT solution, may be of critical importance to encourage adolescents to continue using the solution [64].

Figure 6. Sublevel models for the functional goal "increase awareness." The "*" in the model represents findings from the literature; the "#" represents suggestions by experts.



Other Associated Aspects

To strengthen the effect of ICT solutions, attention should be paid to the user experience and the usefulness of the information provided, as both have been demonstrated to strengthen the effect of the solution [46,69,83]. The need for a positive user experience through a simple and well-ordered design that is pleasant to use has also been highlighted by the users involved in previous studies [43,46,54,60,69].

The opinions of adolescents concerning different gamification elements are contradictory, as some studies with adolescents have found rewarding to be an irrelevant feature [64], whereas some other studies emphasize the importance of rewards [58]. It has been suggested that different gamification elements such as avatars and rewards to achieve a goal or meet a challenge increase the attractiveness of and engagement with an ICT solution and reduce dropout [44,53]. It has also been pointed out that rewards have the potential to foster motivation and increase self-efficacy [46]. However, the logic behind receiving award points must be clear to the participants [55,69], and the reward system should be carefully planned so that it would support intrinsic motivations because intrinsically motivated behaviors are derived from enjoyment and interest by the performer [64,84]. This, in turn, supports long-term behavior changes [85]. Adolescents have pointed out that it is not just the goal achievement that should be rewarded but also the improvements and progress toward goal achievement [58].

It is widely accepted that the learning process should support the autonomy of a child, as it affects the achievement and motivation of the child [84,86]. The goal models proposed by us include elements of support for all three categories of autonomy: organizational, procedural, and cognitive. Support for organizational autonomy involves students in decision making on management issues [84], which in our goal models is expressed by the possibilities to choose group members for challenges or set dates for achievement goals. Support for procedural autonomy enables one to choose their own way of finding a solution [84], which is in our model addressed by, for example, providing information about different PA opportunities and challenges that the adolescent can choose from to achieve their goals. Support for cognitive autonomy encourages student ownership of the learning process and is considered the most important type of autonomy to be supported in the learning process [84]. Some examples of the features supporting cognitive autonomy included by our goal models are providing

XSL•FO
RenderX

feedback, enabling the formulation of personal goals, and providing means to reach one's goals independently.

Elaboration to User Stories

In this subsection, we describe how goal models can be further elaborated to more detailed technical requirements for an ICT solution in the form of user stories [39], which is one of the most prevalent ways of representing requirements in agile software engineering [73,74]. Table 1 presents user stories associated with the submodel *Participate in challenges* shown in Figure 4, following the user story format adapted from Cohn [73], as follows: "As a [user performing a certain role], I need [to perform action] to support [achieving a certain goal]".

Table 1. User stories for the submodel "Participate in challenges."

Role and action	Goal
Adolescent	
Filter by intensity	To find challenge
Filter by duration	To find challenge
Filter by aim	To find challenge
Filter by rating	To find challenge
Invite friends to participate in challenges	To create group
View own progress	To view results
View comparison with friends	To view results
View comparison with selected groups	To view results
Coach or teacher	
Filter by intensity	To find challenge
Filter by duration	To find challenge
Filter by aim	To find challenge
Filter by rating	To find challenge
Invite friends to participate in challenges	To create group
View comparison with friends	To view results
View comparison with selected groups	To view results

Feedback From Experts

The experts considered an agent-oriented goal model to be an effective representation artifact to present the features that support the PA of adolescents and facilitate an understanding of the problem domain. The models created in our study also proved to be effective communication tools because expressing the existing scientific evidence by means of clear and simple goal models, together with just short explanations, helped the experts to understand the ideas expressed by the models and focus on analyzing and enhancing them. The created goal models provide a good overview of the necessary features of ICT solutions. The experts pointed out that although the notation of the models was easy to understand, to fully understand the presented ideas, additional information provided by the researcher was useful. It was also highlighted that the goal models created in the process could serve as input for developing ICT-based interventions because goal models presented all important and evidence-based factors in a holistic and systematic way. This would make the agent-oriented goal modeling an easily usable method for selecting and planning further intervention activities.

Discussion

Principal Findings

This study proposes a novel approach to create agent-oriented goal models by combining a literature review and qualitative information from experts. According to our experience, this approach is a promising and feasible method for identifying the functional, quality, and emotional goals of ICT solutions for supporting PA in adolescents. The literature review conducted by us as the first step enabled the collection of existing scientific evidence and expressed it through holistic goal models capable of representing functional, quality, and emotional aspects and relationships between them. The resulting goal models helped to obtain and present an overview of the existing problems and possible solutions and easily communicate them to nontechnical experts. As a result, the experts could focus on the aspects that were missing from the models, and they did not have to spend time or effort to point out the existing theories or previously identified possibly effective features. Such an approach also saved experts a lot of time, which is an important aspect considering the need to retain fruitful cooperation with experts and therefore cannot be underestimated. As most of the experts involved had previous experience in implementing PA interventions in a school setting or practical work experience from school, they could focus on adding valuable knowledge

XSL•FO
RenderX

from the field and pointing out aspects that have a high potential to support the usability and effectiveness of ICT solutions in the target group. Examples and experiences from real life provided by experts constitute invaluable information to better understand the existing problems and possible solutions.

Several goals included by the goal models emerging from the study indicate that ICT solutions aimed at supporting PA can benefit greatly from applying machine learning methods. For example, machine learning methods can contribute to a better user experience by providing tailored advice and information. Moreover, well-timed support by the ICT solution can reduce the odds of exercise relapse or dropout [87]. In addition, machine learning algorithms can help to adjust personalized goals [87] or form adaptive goals that, according to randomized controlled trials with adults [88,89], are more effective than static goals. Therefore, future ICT developments should consider applying different machine learning methods to support the PA of users.

In this study, we preferred to conduct one-on-one interviews with experts instead of workshops or focus groups, which were often used in previous studies to identify the functional, quality, and emotional goals to be represented by goal models [30,32]. The main reason for this decision was the COVID-19 situation, which has set restrictions on physical indoor meetings in larger groups. Therefore, the interviews were conducted on the web or outdoors, where the risk of infection was lower. Such an approach certainly has its advantages and disadvantages. During one-on-one interviews, the experts were able to freely express their opinions and ideas without being interrupted, having additional discussions, or experiencing the need to further justify the ideas presented by them. In addition, the situation created by COVID-19 enabled us to refrain from *reinventing the wheel*, as the starting point of the discussions consisted of goal models created by the researcher based on a thorough literature review. On the other hand, workshops support the cocreation process in which an idea or example by one stakeholder can be developed further by the other stakeholders. The research conducted by us was a pilot study, and further studies should validate our findings and improve the goal models with a wider range of stakeholders, such as parents; coaches; and naturally, the main target group—adolescents. However, we believe that the models included in this study lay a strong foundation for further development.

Conclusions

This paper presents the first attempt to create agent-oriented goal models of ICT solutions that support the PA of adolescents by combining evidence-based information and the opinions of experts and practitioners. The desired ICT solutions have the potential to contribute to the fight against low PA levels among adolescents by including in ICT solutions the features supported by the evidence. The strength of the proposed model is, in addition to including information from the scientific literature and experts in the field, a holistic and coherent presentation of early functional, quality, and emotional requirements. Our approach seems to be efficient in communicating the results of literature reviews to experts and supporting collaboration with nontechnical stakeholders. We believe that the created models have a high potential to help requirements engineers and developers to provide more efficient ICT solutions in the future.

Acknowledgments

The authors would like to thank all the experts who provided invaluable input to the models. This research was supported by the *Pilots for Healthy and Active Ageing* (Pharaon) project (1.12.2019-30.11.2023) funded by the European Union's Horizon 2020 research and innovation program under grant 857188.

Conflicts of Interest

None declared.

References

1. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
2. Poitras VJ, Gray CE, Borghese MM, Carson V, Chaput J, Janssen I, et al. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):197-239. [doi: 10.1139/apnm-2015-0663] [Medline: 27306431]
3. Strong WB, Malina RM, Blimkie CJ, Daniels SR, Dishman RK, Gutin B, et al. Evidence based physical activity for school-age youth. J Pediatr 2005 Jun;146(6):732-737. [doi: 10.1016/j.jpeds.2005.01.055] [Medline: 15973308]
4. Kalman M, Inchley J, Sigmundova D, Iannotti R, Tynjälä JA, Hamrik Z, et al. Secular trends in moderate-to-vigorous physical activity in 32 countries from 2002 to 2010: a cross-national perspective. Eur J Public Health 2015 Apr;25 Suppl 2:37-40. [doi: 10.1093/eurpub/ckv024] [Medline: 25805785]
5. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva: World Health Organisation; 2010.
6. Dumith SC, Gigante DP, Domingues MR, Kohl HW. Physical activity change during adolescence: a systematic review and a pooled analysis. Int J Epidemiol 2011 Jun 18;40(3):685-698. [doi: 10.1093/ije/dyq272] [Medline: 21245072]

XSL•FO
RenderX

7. Collings P, Wijndaele K, Corder K, Westgate K, Ridgway C, Sharp S, et al. Magnitude and determinants of change in objectively-measured physical activity, sedentary time and sleep duration from ages 15 to 17.5y in UK adolescents: the ROOTS study. Int J Behav Nutr Phys Act 2015 May 14;12:61 [FREE Full text] [doi: 10.1186/s12966-015-0222-4] [Medline: 25971606]
8. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195. [doi: 10.1159/000222244] [Medline: 20054224]
9. Sawyer SM, Afifi RA, Bearinger LH, Blakemore S, Dick B, Ezeh AC, et al. Adolescence: a foundation for future health. Lancet 2012 Apr;379(9826):1630-1640. [doi: 10.1016/S0140-6736(12)60072-5]
10. Mannocci A, D'Egidio V, Backhaus I, Federici A, Sinopoli A, Varela AR, et al. Are there effective interventions to increase physical activity in children and young people? An umbrella review. Int J Environ Res Public Health 2020 May 18;17(10):3528 [FREE Full text] [doi: 10.3390/ijerph17103528] [Medline: 32443505]
11. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
12. Carlin A, Murphy MH, Gallagher AM. Current influences and approaches to promote future physical activity in 11-13 year olds: a focus group study. BMC Public Health 2015 Dec 21;15(1):1270 [FREE Full text] [doi: 10.1186/s12889-015-2601-9] [Medline: 26689932]
13. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
14. Hamel LM, Robbins LB. Computer- and web-based interventions to promote healthy eating among children and adolescents: a systematic review. J Adv Nurs 2013 Jan 04;69(1):16-30. [doi: 10.1111/j.1365-2648.2012.06086.x] [Medline: 22757605]
15. McIntosh JR, Jay S, Hadden N, Whittaker PJ. Do e-Health interventions improve physical activity in young people: a systematic review. Public Health 2017 Jul;148:140-148. [doi: 10.1016/j.puhe.2017.04.001] [Medline: 28482322]
16. 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]
17. Chen J, Wilkosz ME. Efficacy of technology-based interventions for obesity prevention in adolescents: a systematic review. Adolesc Health Med Ther 2014 Aug:159. [doi: 10.2147/ahmt.s39969]
18. Rose T, Barker M, Jacob CM, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
19. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
20. Shin Y, Kim SK, Lee M. Mobile phone interventions to improve adolescents' physical health: a systematic review and meta-analysis. Public Health Nurs 2019 Nov 28;36(6):787-799. [doi: 10.1111/phn.12655] [Medline: 31456259]
21. Lau PW, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
22. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
23. Pope L, Garnett B, Dibble M. Lessons learned through the implementation of an eHealth physical activity gaming intervention with high school youth. Games Health J 2018 Apr;7(2):136-142. [doi: 10.1089/g4h.2017.0164] [Medline: 29393679]
24. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May 13;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
25. Schoffman DE, Turner-McGrievy G, Jones SJ, Wilcox S. Mobile apps for pediatric obesity prevention and treatment, healthy eating, and physical activity promotion: just fun and games? Transl Behav Med 2013 Sep;3(3):320-325 [FREE Full text] [doi: 10.1007/s13142-013-0206-3] [Medline: 24073184]
26. Mendoza A, Miller T, Pedell S, Sterling L. The role of users' emotions and associated quality goals on appropriation of systems: two case studies. In: Proceedings of the 24th Australasian Conference on Information Systems. 2013 Presented at: 24th Australasian Conference on Information Systems; Dec 4-6, 2013; Melbourne p. 1-11 URL: https://people.eng.unimelb.edu.au/tmiller/pubs/acis13.pdf
27. Slootmaker SM, Chinapaw MJ, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]

XSL•FO
RenderX

28. Curumsing MK, Fernando N, Abdelrazek M, Vasa R, Mouzakis K, Grundy J. Emotion-oriented requirements engineering: a case study in developing a smart home system for the elderly. J Syst Softw 2019 Jan;147:215-229. [doi: 10.1016/j.jss.2018.06.077]
29. Miller T, Pedell S, Sterling L, Vetere F, Howard S. Understanding socially oriented roles and goals through motivational modelling. J Syst Softw 2012 Sep;85(9):2160-2170. [doi: 10.1016/j.jss.2012.04.049]
30. Taveter K, Sterling L, Pedell S, Burrows R, Taveter E. A method for eliciting and representing emotional requirements: two case studies in e-Healthcare. In: Proceedings of the IEEE 27th International Requirements Engineering Conference Workshops (REW). 2019 Presented at: 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW); Sept. 23-27, 2019; Jeju, Korea (South). [doi: 10.1109/rew.2019.00021]
31. Sterling LS, Taveter K. The Art of Agent-Oriented Modeling. Cambridge: MIT Press; 2009:1-978.
32. Lorca A, Burrows R, Sterling L. Teaching motivational models in agile requirements engineering. In: Proceedings of the IEEE 8th International Workshop on Requirements Engineering Education and Training (REET). 2018 Presented at: IEEE 8th International Workshop on Requirements Engineering Education and Training (REET); Aug. 21, 2018; Banff, AB, Canada. [doi: 10.1109/reet.2018.00010]
33. Baxter G, Sommerville I. Socio-technical systems: from design methods to systems engineering. Interact Comput 2011 Jan;23(1):4-17. [doi: 10.1016/j.intcom.2010.07.003] [Medline: 25904163]
34. Sommerville I. Software Engineering. 9th ed. London, UK: Pearson; 2010:1-792.
35. Miller T, Lu B, Sterling L, Beydoun G, Taveter K. Requirements elicitation and specification using the agent paradigm: the case study of an aircraft turnaround simulator. IIEEE Trans Software Eng 2014 Oct 1;40(10):1007-1024. [doi: 10.1109/tse.2014.2339827]
36. Miller T, Pedell S, Lopez-Lorca AA, Mendoza A, Sterling L, Keirnan A. Emotion-led modelling for people-oriented requirements engineering: the case study of emergency systems. J Syst Softw 2015 Jul;105:54-71. [doi: 10.1016/j.jss.2015.03.044]
37. Shvartsman I, Taveter K, Parmak M, Meriste M. Agent-oriented modelling for simulation of complex environments. In: Proceedings of the International Multiconference on Computer Science and Information Technology. 2010 Presented at: International Multiconference on Computer Science and Information Technology; Oct. 18-20, 2010; Wisla, Poland. [doi: 10.1109/imcsit.2010.5679731]
38. Shvartsman I, Taveter K. Agent-oriented knowledge elicitation for modeling the winning of "hearts and minds". In: Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS). 2011 Presented at: Federated Conference on Computer Science and Information Systems (FedCSIS); Sept. 18-21, 2011; Szczecin, Poland.
39. Tenso T, Norta A, Rootsi H, Taveter K, Vorontsova I. Enhancing requirements engineering in agile methodologies by agent-oriented goal models: two empirical case studies. In: Proceedings of the IEEE 25th International Requirements Engineering Conference Workshops (REW). 2017 Presented at: IEEE 25th International Requirements Engineering Conference Workshops (REW); Sept. 4-8, 2017; Lisbon, Portugal. [doi: 10.1109/rew.2017.24]
40. Quelly SB, Norris AE, DiPietro JL. Impact of mobile apps to combat obesity in children and adolescents: a systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17. [doi: 10.1111/jspn.12134] [Medline: 26494019]
41. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]
42. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. Soc Sci Med 2017 Nov;192:125-133. [doi: 10.1016/j.socscimed.2017.09.021] [Medline: 28965003]
43. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
44. Monteiro-Guerra F, Rivera-Romero O, Fernandez-Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2020 Jun;24(6):1738-1751. [doi: 10.1109/jbhi.2019.2947243]
45. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 05;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
46. Hosseinpour M, Terlutter R. Your personal motivator is with you: a systematic review of mobile phone applications aiming at increasing physical activity. Sports Med 2019 Sep;49(9):1425-1447 [FREE Full text] [doi: 10.1007/s40279-019-01128-3] [Medline: 31144235]
47. Casey MM, Eime RM, Payne WR, Harvey JT. Using a socioecological approach to examine participation in sport and physical activity among rural adolescent girls. Qual Health Res 2009 Jul 25;19(7):881-893. [doi: 10.1177/1049732309338198] [Medline: 19556398]

XSL•FO
RenderX

48. Jago R, Macdonald-Wallis K, Thompson JL, Page AS, Brockman R, Fox KR. Better with a buddy: influence of best friends on children's physical activity. Med Sci Sports Exerc 2011 Feb;43(2):259-265. [doi: 10.1249/MSS.0b013e3181edefaa] [Medline: 20581714]
49. Wilson D, Williams J, Evans A, Mixon G, Rheaume C. Brief report: a qualitative study of gender preferences and motivational factors for physical activity in underserved adolescents. J Pediatr Psychol 2005;30(3):293-297. [doi: 10.1093/jpepsy/jsi039]
50. Martins J, Marques A, Sarmento H, Carreiro da Costa F. Adolescents' perspectives on the barriers and facilitators of physical activity: a systematic review of qualitative studies. Health Educ Res 2015 Oct 31;30(5):742-755. [doi: 10.1093/her/cyv042] [Medline: 26324394]
51. 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) [FREE Full text] [Medline: 27034992]
52. Goodyear VA, Kerner C, Quennerstedt M. Young people's uses of wearable healthy lifestyle technologies; surveillance, self-surveillance and resistance. Sport Educ Soc 2017 Sep 22;24(3):212-225. [doi: 10.1080/13573322.2017.1375907]
53. Corepal R, Best P, O'Neill R, Tully MA, Edwards M, Jago R, et al. Exploring the use of a gamified intervention for encouraging physical activity in adolescents: a qualitative longitudinal study in Northern Ireland. BMJ Open 2018 Apr 20;8(4):e019663 [FREE Full text] [doi: 10.1136/bmjopen-2017-019663] [Medline: 29678971]
54. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
55. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
56. Brooks F, Magnusson J. Taking part counts: adolescents' experiences of the transition from inactivity to active participation in school-based physical education. Health Educ Res 2006 Dec 01;21(6):872-883. [doi: 10.1093/her/cyl006] [Medline: 16714367]
57. Moore JB, Jilcott SB, Shores KA, Evenson KR, Brownson RC, Novick LF. A qualitative examination of perceived barriers and facilitators of physical activity for urban and rural youth. Health Educ Res 2010 Apr 18;25(2):355-367. [doi: 10.1093/her/cyq004] [Medline: 20167607]
58. Pope L, Garnett B, Dibble M. Engaging adolescents to inform the development of a mobile gaming app to incentivize physical activity. JMIR Res Protoc 2017 Aug 29;6(8):e161 [FREE Full text] [doi: 10.2196/resprot.8113] [Medline: 28851676]
59. den Akker HO, Cabrita M, den Akker RO, Jones VM, Hermens HJ. Tailored motivational message generation: a model and practical framework for real-time physical activity coaching. J Biomed Inform 2015 Jun;55:104-115 [FREE Full text] [doi: 10.1016/j.jbi.2015.03.005] [Medline: 25843359]
60. Tong H, Coiera E, Laranjo L. Using a mobile social networking app to promote physical activity: a qualitative study of users' perspectives. J Med Internet Res 2018;20(12):e11439. [doi: 10.2196/11439]
61. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
62. 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]
63. Shameli A, Althoff T, Saberi A, Leskovec J. How gamification affects physical activity: large-scale analysis of walking challenges in a mobile application. In: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: WWW '17: 26th International World Wide Web Conference; April, 2017; Perth Australia p. 455-463. [doi: 10.1145/3041021.3054172]
64. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
65. Drehlich M, Naraine M, Rowe K, Lai SK, Salmon J, Brown H, et al. Using the technology acceptance model to explore adolescents' perspectives on combining technologies for physical activity promotion within an intervention: usability study. J Med Internet Res 2020 Mar 06;22(3):e15552 [FREE Full text] [doi: 10.2196/15552] [Medline: 32141834]
66. Humbert ML, Chad KE, Spink KS, Muhajarine N, Anderson KD, Bruner MW, et al. Factors that influence physical activity participation among high- and low-SES youth. Qual Health Res 2006 Apr;16(4):467-483. [doi: 10.1177/1049732305286051] [Medline: 16513991]
67. Finnerty T, Reeves S, Dabinett J, Jeanes YM, Vögele C. Effects of peer influence on dietary intake and physical activity in schoolchildren. Public Health Nutr 2009 Sep 01;13(3):376-383. [doi: 10.1017/s1368980009991315]
68. Salvy S, Roemmich JN, Bowker JC, Romero ND, Stadler PJ, Epstein LH. Effect of peers and friends on youth physical activity and motivation to be physically active. J Pediatr Psychol 2009 Mar 24;34(2):217-225 [FREE Full text] [doi: 10.1093/jpepsy/jsn071] [Medline: 18617572]

XSL•FO
RenderX

69. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
70. Dishman RK, Motl RW, Saunders R, Felton G, Ward DS, Dowda M, et al. Enjoyment mediates effects of a school-based physical-activity intervention. Med Sci Sports Exerc 2005 Mar;37(3):478-487. [doi: 10.1249/01.mss.0000155391.62733.a7] [Medline: 15741848]
71. Dishman R, Sallis J, Orenstein D. The determinants of physical activity and exercise. Public Health Rep 1985;100(2):158-171 [FREE Full text] [Medline: 3920714]
72. Jago R, Brockman R, Fox KR, Cartwright K, Page AS, Thompson JL. Friendship groups and physical activity: qualitative findings on how physical activity is initiated and maintained among 10-11 year old children. Int J Behav Nutr Phys Act 2009 Jan 12;6(1):4 [FREE Full text] [doi: 10.1186/1479-5868-6-4] [Medline: 19138411]
73. Cohn M. User Stories Applied: For Agile Software Development. Boston, Massachusetts, United States: Addison-Wesley Professional; 2004:1-304.
74. Paetsch F, Eberlein A, Maurer F. Requirements engineering and agile software development. In: Proceedings of the Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003. 2003 Presented at: Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003; June 11, 2003; Linz, Austria. [doi: 10.1109/enabl.2003.1231428]
75. Vallgårda S. Nudge: a new and better way to improve health? Health Policy 2012 Feb;104(2):200-203. [doi: 10.1016/j.healthpol.2011.10.013] [Medline: 22113151]
76. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
77. van Genugten G, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. J Med Internet Res 2016;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
78. Gaudet J, Gallant F, Bélanger M. A fit of bit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
79. Dahlgren G, Whitehead M. Policies and Strategies to Promote Social Equity in Health. Stockholm, Sweden: Institute for Futures Studies; 1991.
80. Gilchrist P, Wheaton B. The social benefits of informal and lifestyle sports: a research agenda. Int J Sport Policy Polit 2017;9(1):1-10. [doi: 10.1080/19406940.2017.1293132]
81. Sanders T, Feng X, Fahey PP, Lonsdale C, Astell-Burt T. The influence of neighbourhood green space on children's physical activity and screen time: findings from the longitudinal study of Australian children. Int J Behav Nutr Phys Act 2015 Sep 30;12(1):126 [FREE Full text] [doi: 10.1186/s12966-015-0288-z] [Medline: 26419752]
82. Corder K, Schiff A, Kesten JM, van Sluijs EM. Development of a universal approach to increase physical activity among adolescents: the GoActive intervention. BMJ Open 2015 Aug 25;5(8):e008610 [FREE Full text] [doi: 10.1136/bmjopen-2015-008610] [Medline: 26307618]
83. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
84. Stefanou CR, Perencevich KC, DiCintio M, Turner JC. Supporting autonomy in the classroom: ways teachers encourage student decision making and ownership. Educ Psychol 2004 Jun;39(2):97-110. [doi: 10.1207/s15326985ep3902_2]
85. Ahn SJ, Johnsen K, Ball C. Points-based reward systems in gamification impact children's physical activity strategies and psychological needs. Health Educ Behav 2019 Jun 24;46(3):417-425 [FREE Full text] [doi: 10.1177/1090198118818241] [Medline: 30678507]
86. Reeve J, Halusic M. How K-12 teachers can put self-determination theory principles into practice. Theory Res Educ 2009 Jun 25;7(2):145-154. [doi: 10.1177/1477878509104319]
87. Zhou M, Fukuoka Y, Goldberg K, Vittinghoff E, Aswani A. Applying machine learning to predict future adherence to physical activity programs. BMC Med Inform Decis Mak 2019 Aug 22;19(1):169 [FREE Full text] [doi: 10.1186/s12911-019-0890-0] [Medline: 31438926]
88. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013 Dec;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
89. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]

Abbreviations

BCT: behavior change technique

XSL•FO
RenderX

ICT: information and communication technology
PA: physical activity
STS: sociotechnical system

Edited by G Eysenbach; submitted 06.10.20; peer-reviewed by K Parker, KT Win; comments to author 18.11.20; revised version received 28.12.20; accepted 02.04.21; published 19.05.21.

Please cite as:
Mooses K, Taveter K
Agent-Oriented Goal Models in Developing Information Systems Supporting Physical Activity Among Adolescents: Literature Review and Expert Interviews
J Med Internet Res 2021;23(5):e24810
URL: https://www.jmir.org/2021/5/e24810
doi:10.2196/24810
PMID:34009127

©Kerli Mooses, Kuldar Taveter. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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 Internet-Based Cognitive Behavioral Therapy for Insomnia: Health-Economic Evaluation From the Societal and Public Health Care Perspective Alongside a Randomized Controlled Trial

Claudia Buntrock[1], PhD; Dirk Lehr[2], PhD; Filip Smit[3,4,5], PhD; Hanne Horvath[6], PhD; Matthias Berking[1], PhD; Kai Spiegelhalder[7], PhD; Heleen Riper[5,8], PhD; David Daniel Ebert[9], PhD

[1]Chair of Clinical Psychology and Psychotherapy, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

[2]Institute of Psychology, Leuphana University Lueneburg, Lueneburg, Germany

[3]Center for Economic Evaluation and Machine Learning, Trimbos Insitute, Utrecht, Netherlands

[4]Department of Biostatistics and Epidemiology, Amsterdam Public Health Research Institute, VU University Medical Center Amsterdam, Amsterdam, Netherlands

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

[6]GET.ON Institute, Hamburg, Germany

[7]Department of Psychiatry and Psychotherapy, Medical Center University of Freiburg, Freiburg, Germany

[8]Telepsychiatric Centre, University of Southern Denmark, Odense, Denmark

[9]Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany

Corresponding Author:
Claudia Buntrock, PhD
Chair of Clinical Psychology and Psychotherapy
Friedrich-Alexander University Erlangen-Nuremberg
Nägelsbachstr 25a
Erlangen, 91052
Germany
Phone: 49 09131 8567568
Email: claudia.buntrock@fau.de

Abstract

Background: The evidence base for internet-based cognitive behavioral therapy for insomnia (iCBT-I) is firm; however, little is known about iCBT-I's health-economic effects.

Objective: This study aimed to evaluate the cost-effectiveness and cost–utility of iCBT-I in reducing insomnia among schoolteachers.

Methods: Schoolteachers (N=128) with clinically significant insomnia symptoms and work-related rumination were randomized to guided iCBT-I or a wait list control group, both with unrestricted access to treatment as usual. Health care use, patient and family expenditures, and productivity losses were self-assessed and used for costing from a societal and a public health care perspective. Costs were related to symptom-free status (score <8 on the insomnia severity index) and quality-adjusted life years (QALYs) gained. Sampling error was handled using nonparametric bootstrapping.

Results: Statistically significant differences favoring the intervention group were found for both health outcomes (symptom-free status yes or no: β=.30; 95% CI 0.16-0.43; QALYs: β=.019, 95% CI 0.01-0.03). From a societal perspective, iCBT-I had a 94% probability of dominating the wait list control for both health outcomes. From a public health care perspective, iCBT-I was more effective but also more expensive than the wait list control, resulting in an incremental cost-effectiveness ratio of €650 per symptom-free individual. In terms of QALYs, the incremental cost-effectiveness ratio was €11,285. At a willingness-to-pay threshold of €20,000 per QALY gained, the intervention's probability of being cost-effective was 89%.

Conclusions: Our trial indicates that iCBT could be considered as a good value-for-money intervention for insomnia.

Trial Registration: German Clinical Trial Registry: DRKS00004700; https://tinyurl.com/2nnk57jm

International Registered Report Identifier (IRRID): RR2-10.1186/1745-6215-14-169

(J Med Internet Res 2021;23(5):e25609) doi:10.2196/25609

XSL•FO
RenderX

KEYWORDS

insomnia; internet-based cognitive behavioural therapy; iCBT-I; economic evaluation; cost-effectiveness; cost-utility; cognitive behavioral therapy; behavior; sleep; economics; public health; perspective

Introduction

Insomnia is characterized by difficulties in initiating or maintaining sleep and/or early morning awakenings that occur 3 nights or more per week, for at least 3 months, resulting in poor sleep quality and significant daytime impairment [1]. Insomnia is one of the most common sleep disorders among adults. Prevalence estimates range from 6% for insomnia disorder [2] to 25% for insomnia symptoms [3].

Insomnia is associated with a range of adverse health consequences for individuals, including poor daytime functioning [4] and reduced health-related quality of life [5]. In view of its high prevalence and its debilitating nature, insomnia is related to a substantial health and economic burden. As such, it increases the risk of future mental disorders (eg, major depressive disorder) [6]. Economic costs stem from both absenteeism and reduced productivity while at work (ie, presenteeism) [7] as well as increased health care utilization, (eg, medication prescription [8,9]).

In order to reduce the personal and economic burden of insomnia, it is essential to implement interventions that can improve sleep. American and European guidelines recommend cognitive behavioral therapy (CBT) as first-line treatment for insomnia disorder [10,11] due to its substantial clinical evidence base [12]. However, despite this recommendation, CBT for insomnia (CBT-I) is not widely available, mainly due to a shortage of therapists and available resources [13]. In addition, only 37% of those suffering from insomnia seek professional help [3].

Internet-based CBT intervention for insomnia (iCBT-I) has been touted as a solution that can bridge this gap in health care [14]. Meta-analytic evidence demonstrated that iCBT-I is effective in treating insomnia with large effect sizes at posttreatment (eg, Cohen's d = 1.09; 95% CI 0.74-1.45) for insomnia severity [15]. Effect sizes are comparable to those found in individual face-to-face delivered CBT-I (eg, d=1.11; 95% CI 0.94-1.28) [12].

Although the effectiveness of (i)CBT-I has been demonstrated, research on its economic costs and benefits is still limited [16]. Previous reviews have argued that treating insomnia costs less compared to doing nothing [17,18]. A recent review on CBT-I interventions (N=7) using a dominance ranking framework showed that CBT-I was cost-effective compared to pharmacotherapy or no treatment [19]. However, only 2 studies have evaluated the economic effects of an iCBT-I intervention, but they suggested that iCBT-I provides superior health improvements at reduced costs [20,21]. To the best of our knowledge, no study has yet investigated the economic merits of iCBT-I from a societal perspective (including reductions in direct medical costs, patient and family costs, and indirect costs stemming from productivity losses) and from the public health care perspective (including only direct medical costs).

The aim of the this paper was thus to assess, from a societal and public health care perspective, the cost-effectiveness and cost–utility of a guided iCBT-I intervention to reduce insomnia symptoms in currently employed schoolteachers. The health economic evaluation presented here was conducted alongside a randomized controlled trial [22]. A previous publication reported the clinical effects of the iCBT-I intervention (6-month follow-up: Cohen d=1.43; 95% CI 1.04-1.82) [23].

Methods

Study Design

We conducted and reported the health-economic evaluation in agreement with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [24] and the guidelines from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [25]. The economic evaluation was performed from a societal perspective (ie, all relevant costs) and a public health care perspective (ie, direct medical costs) alongside a pragmatic 2-armed randomized controlled trial to establish the cost-effectiveness and cost–utility of a guided iCBT-I intervention as an adjunct to treatment as usual (TAU) for schoolteachers with insomnia compared to a wait list control condition with unrestricted access to TAU. Self-report questionnaires to assess costs and effects were collected at baseline, posttreatment (only health effects; 8 weeks after randomization), and 6-month follow-up via a secured online-based assessment system (Advanced Encryption Standard, 256-bit encrypted). Full details of the trial design can be found elsewhere [22]. The study was approved by the ethics committee of the University of Marburg (reference number: 2013-01K) and is registered under DRKS00004700 in the German clinical trial registry.

Procedure

Participants were recruited in Germany from March 2013 to September 2013 using email distribution lists to primary, secondary, and vocational schools, which were provided by the Ministry of Education in the German state of Nordrhein–Westfalen. Currently employed schoolteachers aged 18 and above with clinically significant insomnia symptoms (Insomnia Severity Index [ISI] >14) and elevated work-related rumination (Irritation scale, subscale "Cognitive Irritation" >14) were included in the study [22]. The exclusion criterion was current psychotherapy for insomnia and/or suicidal ideation (Beck Depression Inventory item on suicidality >1). People taking sleep medication were not excluded from the study but were required to keep their medication on a stable dose during the study period. The Consolidated Standards of Reporting Trials (CONSORT) study flowchart and participants' characteristics at baseline can be found elsewhere [23]. In brief, 128 schoolteachers were recruited into the trial with 64 randomized to the intervention and 64 to the wait list control condition. The average participant was female (95/128, 74%),

XSL•FO
RenderX

48 years of age (SD 10), married or in a partnership (92/128, 72%), and had a diagnosis of a primary insomnia (100/128, 78.1%) with moderate severity, and 14% (18/128) had a comorbid major depression [23]. Randomization took place at the individual level in a ratio of 1:1 and was conducted centrally by an independent research staff member not otherwise involved in the study using an automated web-based program (randomisation.eu). Study participants were not masked to their treatment allocation due to the nature of the psychological intervention.

Interventions

In this pragmatic trial, all study participants had unrestricted access to TAU. In Germany, TAU for elevated insomnia symptoms usually includes visits to the general practitioner followed by more intensive interventions, such as cognitive behavioral therapy and medication if insomnia symptoms persist or worsen.

iCBT-I Intervention (GET.ON Recovery)

The iCBT-I intervention (GET.ON Recovery [22,23]) has been specifically tailored to schoolteachers experiencing work-related stress and insomnia. The intervention was mainly based on cognitive behavioral methods (eg, sleep restriction therapy, stimulus control therapy, relaxation, sleep hygiene, and cognitive interventions) [26]. These methods were supplemented by techniques effective in reducing work stress and fostering mental detachment from work-related problems derived from behavioral activation [27], metacognitive therapy [28], gratitude research [29], and research on boundary management [30]. The intervention consisted of 6 weekly modules. Overall, of the 64 participants, 61 (95.3%) completed all 6 modules [23]. Participants received written feedback on each completed module by an eCoach (ie, a trained clinical psychologist), who followed a standardized coaching manual. To maximize the comparability of the participants and maintain the guidance at a minimal level, eCoaches were advised that the time spent on each participant per module should not exceed 30 minutes; thus, the total amount of time spent on each participant was approximately 3 hours for the total duration of the intervention [23]. eCoaches were supervised by a clinical psychologist.

Wait List Control Condition

In addition to TAU, individuals in the control group were eventually granted access to the unguided version of the intervention after completing the final assessment at 6 months post baseline.

Outcomes

ISI Symptom-Free Status

The health outcome in the cost-effectiveness analysis was symptom-free status defined as a score <8 on the ISI [31]. The ISI is a 7-item instrument answered on a 5-point Likert scale with a total score ranging from 0 to 28. The psychometric properties of the online version of the questionnaire have been well established [32]. In the current study, internal consistency was set at Cronbach α=.91.

Quality-Adjusted Life Years

Quality-adjusted life years (QALYs) were used as a health outcome in the cost–utility analysis (CUA). QALYs were based on the 6D Health State Short Form (SF-6D; a subset of 6 items of the Short Form Health Survey Version 1 [33]). The SF-6D contains 6 dimensions (each with between 2 and 5 levels) and can generate 7500 different health states. Utility values were derived using Brazier's algorithm [34]. QALY health gains were estimated by calculating the area under the curve of linearly interpolated SF-6D utilities between measurements to cover the whole 6-month follow-up period. The SF-6D was used in this study because the instrument is known to be more sensitive to changes in mild to moderate physical and mental health conditions than is the EQ-5D-3L questionnaire [35,36].

Resource Use and Costing

We used the Trimbos and Institute for Medical Technology Assessment "Treatment Inventory of Costs in Patients with psychiatric disorders" questionnaire (TiC-P) to collect data on health care utilization, patient and family costs, and productivity losses [37,38]. The TiC-P is a retrospective questionnaire with a 3-month recall period. The TiC-P was adapted for use in Germany and has been used in a large number of cost-effectiveness studies [39-41]. Costs were expressed in euros and indexed for the year 2013, the year the study was conducted, based on the German consumer price index (index factor 1.04) [42]. A reference to the National Institute for Health and Care Excellence willingness-to-pay (WTP) threshold of £20,000 (€23,529) to £30 000 (€35,294) per QALY gained was made where appropriate [43]. Costs were converted to pound sterling (£) using the purchasing power parities reported by the Organization for Economic Cooperation and Development. For the reference year 2013, €1 was equated to £0.85.

Intervention Costs

At the time of conducting the study, the market price of the iCBT-I provided by the GET.ON Institute was €299 (£254) per participant including all costs for developing (eg, tailoring intervention content to the target group) and delivering the intervention (eg, eCoaches providing individual feedback to participants).

Health Care Costs

We used 2 German guidelines for calculating health care costs [44,45]. A list of unit cost prices (ie, outpatient care) was used to compute the total health care costs on a per-participant level. Unit cost prices indexed for the year 2013 were as follows: €20.92 (£17.78) for a visit to the general practitioner, €68.06 (£57.85) for an internal medicine consult, €46.55 (£39.57) for a session with a psychiatrist, and €81.44 (£69.22) for a session with a psychotherapist. Hospital stays were computed at €335.52 (£285.19) for an in-patient day in a psychiatric hospital and €306.41 (£260.45) for an in-patient day in a hospital for psychosomatic medicine and psychotherapy. Costs were estimated by multiplying the units of resource use with corresponding unit cost prices.

XSL•FO
RenderX

Medication

The costs of prescribed medication were based on the German drug registry (Rote Liste [46]). Costs of prescribed medication are calculated as the pharmacy retail price, with the pharmacist's "clawback" (ie, wholesale margin) being accounted for. The rates of discount vary between private and statutory (public) health insurances. Therefore, we weighted the mean costs of the 3 largest packages with the same agent based on the daily defined dose by the statutory population share (89% of the German population are statutorily insured).

Patient and Family Costs

Out-of-pocket expenses were directly obtained from participants. Costs for traveling were valued at €0.30 (£0.26) per kilometer for making trips to access health services. Time spent on the intervention and costs of informal care were valued using the opportunity cost method and were estimated at €23.10 (£19.64) per hour [44,45].

Costs of Productivity Losses

Productivity losses can be caused by absenteeism (ie, days not worked) and presenteeism (ie, reduced efficiency while at work). We followed the human capital approach to value costs due to absenteeism [47]. Lost workdays due to absenteeism were valued at the corresponding gross average of participants' income per day. Lost workdays due to presenteeism were computed by taking into account the number of work days for which the participant reported reduced functioning weighted by the reported corresponding inefficiency score for those days (Osterhaus method) [48]. Productivity losses from unpaid work (ie, household work) were valued using the replacement cost method [49]. The estimated value was €18.33 (£15.58) per hour (eg, the average hourly gross wage of domestic help).

Analysis

The study was not powered to statistically test differences in health-economic outcomes. Therefore, we took a probabilistic decision-making approach to make health-economic inferences [50]. We did not discount costs and effects because the analysis was limited to a 6-month time horizon.

In evaluating clinical and cost outcomes, we reported all analyses in accordance with the CONSORT statement [51]. Data were analyzed according to the intention-to-treat principle. Missing data in ISI were imputed via multiple imputation using a Markov Chain Monte Carlo multivariate imputation algorithm (SPSS 21, IBM Corp) with 10 estimations per missing value [23]. To account for missing data in the cost and utility data, we used the regression imputation procedure in Stata version 16 (StataCorp) to obtain the required predicted values [52]. Predictors of outcome and dropout were identified by (logistic) regression analyses. Identifying predictors of outcome helped us to obtain the most likely values of the outcome, whereas identifying predictors of dropout allowed us to correct for bias that might arise by differential loss to follow-up. Differences in effectiveness between study groups at 6-month follow-up were estimated using ordinary least square regression analyses. Due to baseline imbalances, QALYs were adjusted for baseline values [53]. To compare cumulative costs between study groups, a generalized linear regression model with a gamma family distribution and an identity link function was fitted and adjusted for baseline depressive symptom severity and age. The family distribution was selected based on the modified Park test [54]. The identity link function was chosen because an additive effect of the covariates was expected.

Societal costs included all cost categories, while direct medical costs used in analyses from the public health care perspective comprised intervention costs, health care costs, and medication costs. In the cost-effectiveness analysis, the incremental cost-effectiveness ratio (ICER) was calculated by dividing incremental costs (total per-participant costs or direct medical costs) by symptom-free status gained. In the cost–utility analysis, incremental costs were divided by QALYs gained. The corresponding equation was as follows: ICER = (CostsINT – CostsCTR) / (EffectsINT – EffectsCTR), where INT is the intervention group and CTR is the control group [47]. Although costs were gamma distributed, the difference of 2 nonnormally distributed variables (eg, incremental costs) followed a remarkably normal distribution. Hence, to handle sampling uncertainty, we bootstrapped the seemingly unrelated regression equations model ("sureg" command in Stata) to generate 2500 simulations of incremental cost and incremental effect pairs while allowing for correlated residuals of the cost and effect equations and adjusting for potential confounders (eg, baseline utilities in the effect equation; age and baseline depressive symptom severity in the cost equation) [55]. Based on the bootstrapped seemingly unrelated regression equation model, bias-corrected and accelerated 95% CIs were obtained for incremental costs and effects. In addition, 95% CIs around ICERs were obtained by the bootstrap acceptability method [56]. The bootstrapped cost and effect pairs were graphically represented on a cost-effectiveness plane with effects along the horizontal axis and costs along the vertical axis [47]. To assess the probability of the intervention being cost-effective at varying WTP thresholds, cost-effectiveness acceptability curves were plotted [57]. Except for the imputation of missing ISI scores, all analyses were performed using Stata version 16 (StataCorp) [52].

Sensitivity Analyses

Three sensitivity analyses were performed. Presenteeism has been previously identified as one of the main cost drivers in insomnia [20]. As there is no gold standard to measure costs due to presenteeism, we used a different approach to assess presenteeism costs in the first sensitivity analysis. Here, we calculated costs due to presenteeism based on the Health and Labor Questionnaire (HLQ) method. The Osterhaus method, used in the main analyses, tends to overestimate costs due to presenteeism because this method concentrates on the work capacity of the individual. In contrast, the HLQ method focuses on production loss that is recoverable and is not yet made up for, thus generating a lower estimate of costs [58]. In a second sensitivity analysis, we varied the costs of the intervention by plus and minus 20% and 50%, respectively, to reflect uncertainties about the actual market price also including a lower price due to scaling effects. Finally, we conducted a "completers-only analysis" based on the data of the participants who completed the 6-month follow-up assessment.

XSL•FO
RenderX

Results

Study Dropout

At posttreatment, 92.2% (118/128) of the participants completed the follow-up questionnaires, whereas at 6-month follow-up, 88.3% (113/128) did. Dropout rates did not differ between the intervention and control conditions ($\chi 2_1$=1.89; *P*=.17). Study dropout was neither associated with baseline insomnia severity nor with any sociodemographic factors (lowest *P* value=.18 for treatment allocation). There were no missing data due to item nonresponse.

Effects

In the intervention group, 27 out of 64 participants (42%) reached a symptom-free status, whereas in the control group, 4 out of 64 participants (6%) were symptom-free at 6-month follow-up [23]. Statistically significant differences favoring the intervention group were found between the intervention and control group in symptom-free status (incremental effect, (Δ[E]=0.30; 95% CI 0.16-0.43). On average, participants in the intervention group gained 0.36 QALYs (95% CI 0.35-0.37) during the study period, while participants in the control condition gained 0.35 QALYs (95% CI 0.34-0.36). Differences in adjusted incremental QALYs were statistically significant (Δ[E]=0.02; 95% CI 0.01-0.03).

Costs

Both groups showed similar total costs within the 3-month recall period before randomization (intervention group: €2902 [£2467], 95% CI €2111-€3693; control group: €3112 [£2645], 95% CI €2321-€3903). Table 1 presents the 6-month accumulated per-participant costs separately for different cost categories by treatment allocation. Mean direct medical costs were higher in the intervention group (€592 [£503]; 95% CI €444-€740) compared to the control condition (€389 [£331]; 95% CI €242-€537), which could largely be explained by the intervention costs that were only involved in the iCBT-I group. In contrast, both patient and family's costs, along with productivity costs, were lower in the intervention group compared to the control group, with costs related to productivity losses having the largest impact on overall societal costs. Mean per-participant total costs accrued over the 6-month follow-up period were €4030 (£3426; 95% CI €3125-€4934) for the intervention group and €5021 (£4268; 95% CI €3394-€6147) for the control condition. Adjusted incremental differences in total costs were in favor of the intervention group (Δ[C]=–€895 [–£761]; 95% CI –€2155 to €364; ie, lower by €895 in the intervention group).

Table 1. Mean cumulative per-participant costs (in €) by condition over a 6-month follow-up period (based on intention-to-treat sample; N=128).

Costs	Intervention group (n=64), mean (95% CI)	Control group, (n=64), mean (95% CI)	Incremental costs, difference (95% CI)
Direct medical costs			
Intervention costs	299[a]	—[b]	299[a]
General practitioner	36 (21 to 51)	51 (36 to 66)	–15 (–36 to 6)
Mental health care	150 (56 to 244)	163 (69 to 257)	–13 (–146 to 120)
Antidepressants	2 (0 to 5)	5 (2 to 8)	–3 (–7 to 1)
Allied health services[c]	105 (34 to 176)	170 (99 to 241)	–65 (–166 to 36)
Patient and family costs			
Informal care	884 (226 to 1541)	1260 (602 to 1918)	–376 (–1306 to 554)
Domestic help	310 (147 to 474)	(132 to 459)	15 (–217 to 246)
Out-of-pocket expenses[d]	45(7 to 84)	73 (35 to 112)	–28 (-82 to 27) [a]
Travel	8 (3 to 13)	15 (10 to 20)	–7 (–14 to 1)
Productivity costs			
Absenteeism	1005 (473 to 1537)	1104 (573 to 1636)	–99 (–851 to 653)
Presenteeism	1185 (747 to 1623)	1883 (1446 to 2321)	–698
Total costs	4030 (2951 to 5108)	5021 (3942 to 6099)	–991 (–2519 to 534)

[a]As intervention costs are fixed, no 95% CI is applicable here.

[b]Not applicable.

[c]Including physiotherapist, massage, occupational therapist, etc.

[d]For example, allied health services without prescription.

XSL•FO
RenderX

Economic Evaluation

Societal Perspective

Table 2 shows the incremental cost, effects, and cost-effectiveness ratios (based on 2500 bootstrap simulations) for the main analyses. Cost-effectiveness analysis revealed that the iCBT-I intervention resulted in more symptom-free individuals (Δ[E]=0.30; 95% CI 0.16-0.43) and that these health gains were achieved at lower costs (Δ[C]=–€1121 [£953]; 95% CI –€3012 to €64). With regard to the cost-effectiveness plane, most of the replicated ICERs (94%) fell in the south–east quadrant, indicating a 94% probability that the intervention would dominate the control condition (Figure 1). Cost–utility analysis revealed similar results compared to the cost-effectiveness analysis (Table 2). Again, most of the bootstrapped cost and effect pairs (94%) fell in the south–east quadrant, indicating the dominance of the intervention over the control condition (Figure 2). When the societal WTP per additional QALY gained was €0, the iCBT-I intervention had a 94% probability of being more cost-effective than the control condition.

Table 2. Results of the main analyses (based on 2500 bootstrap simulations) based on societal and public health care perspectives.

Type of analysis	Incremental costs (in €), mean (95% CI)[a]	Incremental effects, mean (95% CI)[a]	ICER[b], mean (95% CI)	Distribution over the cost-effectiveness plane (%)			
				NEQ[c]	NWQ[d]	SEQ[e]	SWQ[f]
Societal perspective							
CEA[g] (SFS)[h]	–1121 (–3012 to 64)	0.30 (0.16 to 0.43)	dominant[i]	6	—	94	—
CUA[j] (SF-6D QALY)[k]	–1121 (–3012 to 64)	0.0183 (–0.0182 to 0.0185)	dominant	6	—	94	—
Public health care perspective							
CEA (SFS)	189 (–97 to 350)	0.30 (0.16 to 0.43)	650 (–215 to 1652)	94	—	6	—
CUA (SF-6D QALY)	189 (97 to 350)	0.0183 (0.0182 to 0.0185)	11,285 (–1750 to 27,493)	96	—	4	—

[a]95% CIs in this column were bias-corrected and accelerated.

[b]ICER: incremental cost-effectiveness ratio.

[c]NEQ: north–east quadrant.

[d]NWQ: north–west quadrant.

[e]SEQ: south–east quadrant.

[f]SWQ: south–west quadrant.

[g]CEA: cost-effectiveness analysis.

[h]SFS: symptom-free status (0=no, 1=yes).

[i]dominant: The intervention resulted in higher effects at lower costs compared to the control condition.

[j]CUA: cost–utility analysis.

[k]SF-6D QALY: 6D Health State Short Form quality-adjusted life years based on the SF-12.

XSL•FO
RenderX

Figure 1. Scatterplot of 2500 replicates of the incremental cost and effect pairs (eg, symptom-free status) from the societal perspective on the cost-effectiveness plane: internet-based cognitive behavioral therapy versus wait list control condition and cost-effectiveness acceptability curve.

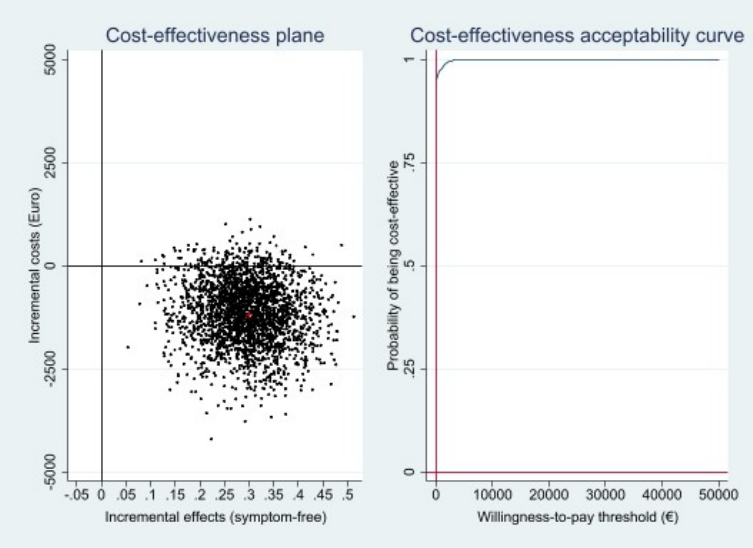
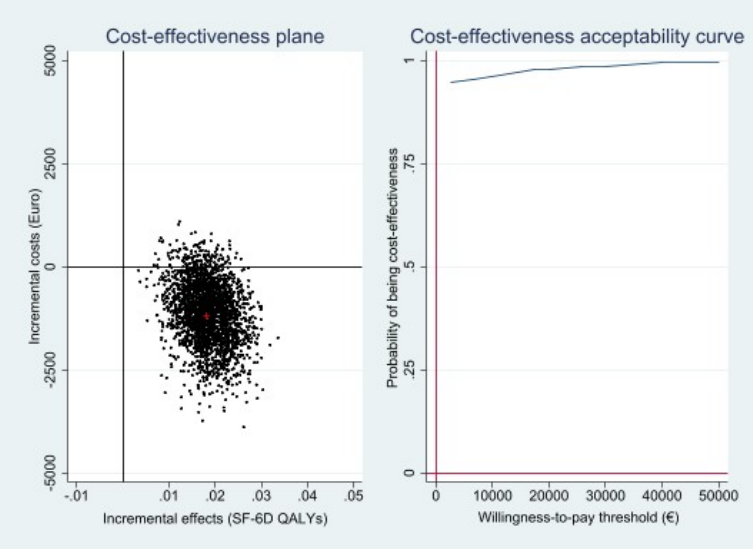


Figure 2. Scatterplot of 2500 replicates of the incremental cost and effect pairs (eg, QALYs gained) from the societal perspective on the cost-effectiveness plane: internet-based cognitive behavioral therapy for insomnia versus wait list control condition and cost-effectiveness acceptability curve. QALY: quality-adjusted life years; SF-6D: 6D Health State Short Form.



Public Health Care Perspective

From a public health care perspective, health benefits were achieved at higher costs (€189 [£161]; 95%CI –€97 to €350). The ICER was €650 (£553; 95% CI –€215 to €1652) for 1 additional symptom-free individual. At a WTP threshold of €0, the iCBT-I intervention's probability of being cost-effective was 6%. With an increase in the WTP to €1500 (£1275) per symptom-free status gained, the probability rose to 96% (Figure 3). Cost–utility analysis revealed an ICER of €11,285 (£9592; 95% CI –€1750 to €27,493) per QALY gained. The corresponding cost-effectiveness acceptability curve shows a probability of 4% and 89% that the intervention would be cost-effective at WTPs of €0 and €20,000 (£17,000) per QALY gained, respectively. With a National Institute for Health and Care Excellence WTP threshold of £20,000 (€23,529) to £30,000 (€35,294) per QALY gained [43], the probability would increase to 95% and 99%, respectively (Figure 4).

XSL•FO
RenderX

Figure 3. Scatterplot of 2500 replicates of the incremental cost and effect pairs (eg, symptom-free status) from the public health care perspective on the cost-effectiveness plane: internet-based cognitive behavioral therapy versus wait list control condition and cost-effectiveness acceptability curve.

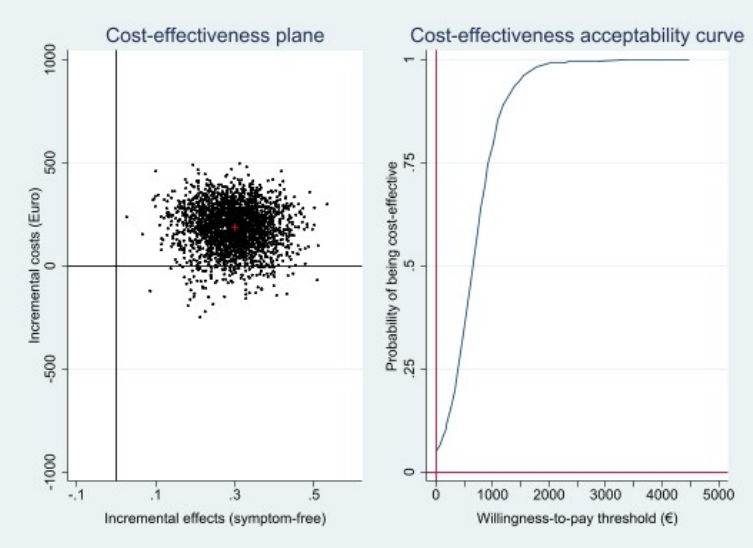
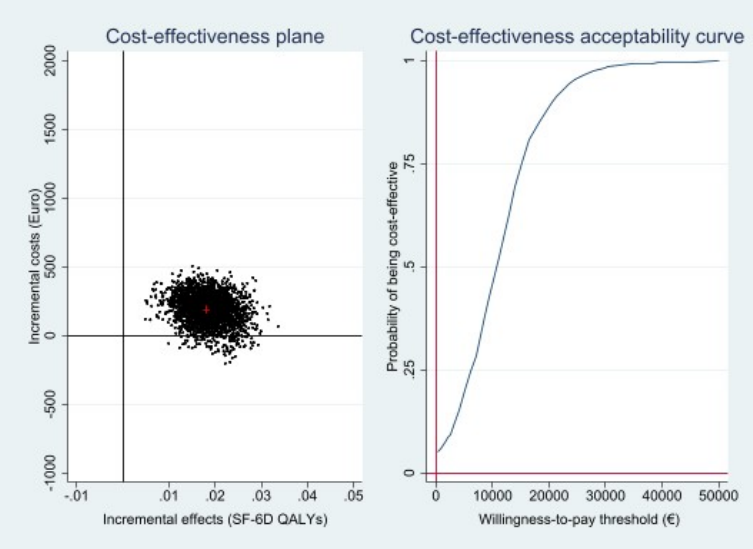


Figure 4. Scatterplot of 2500 replicates of the incremental cost and effect pairs (eg, QALYs gained) from the public health care perspective on the cost-effectiveness plane: internet-based cognitive behavioral therapy versus wait list control condition and cost-effectiveness acceptability curve. QALY: quality-adjusted life years; SF-6D: 6D Health State Short Form.



Sensitivity Analyses

Cost-effectiveness estimates based on the completer-only sample were almost identical to the main analyses, indicating that the imputation procedure did not bias the results (Table 3). From a societal perspective, results of the sensitivity analyses showed that neither using the HLQ method to assess costs due to presenteeism nor increasing intervention costs affected the overall conclusion that the iCBT-I intervention produces greater health gains at lower costs compared with a wait list control condition (Table 3). Using the HLQ method, the intervention's probability of being cost-effective was 72% at a WTP of €0 for both health outcomes. This probability rose to 86% at a WTP of €20,000 (£17,000) per QALY gained (Multimedia Appendix 1 Figure S1). Increasing or decreasing intervention costs by 20% or 50% did not affect the intervention's probability of being cost-effective compared with the main analysis (Table 3). From a public health care perspective, with intervention costs decreased up to 50%, health effects were still gained at slightly higher costs. Increasing the intervention costs by 20% and 50% resulted in an ICER of €14,380 (£12,223) and €19,360 (£16,456) per QALY gained, respectively (Table 3). At a WTP of €20,000 per QALY gained, the intervention's probability of being cost-effective was 80% and 60% (Multimedia Appendix 1 Figure S2), respectively.

XSL•FO
RenderX

Table 3. Results of the sensitivity analyses (based on 2500 bootstrap simulations) based on the societal and public health care perspective.

Type of analysis	Incremental costs (in €), mean (95% CI)[a]	Incremental effects, mean (95% CI)[a]	ICER[b], mean (95% CI)	Distribution over the cost-effectiveness plane (%)			
				NEQ[c]	NWQ[d]	SEQ[e]	SWQ[f]
Societal perspective							
Presenteeism costs based on HLQ method[g]							
CEA[h]	–459 (–2155 to 796)	0.30 (0.16 to 0.43)	dominant[i]	28	—	72	—
CUA[j]	–459 (–2155 to 796)	0.0183 (0.0182 to 0.0185)	dominant	28	—	72	—
Intervention costs plus 20%							
CEA	–1062 (–2801 to 192)	0.30 (0.16 to 0.43)	dominant	6	—	94	—
CUA	–1062 (–2801 to 192)	0.0183 (0.0182 to 0.0185)	dominant	7	—	93	—
Intervention costs plus 50%							
CEA	–972 (–2804 to 249)	0.30 (0.16 to 0.43)	dominant	8	—	92	—
CUA	–972 (–2804 to 249)	0.0183 (.0182 to 0.0185)	dominant	9	—	91	—
Intervention costs minus 20%							
CEA	–1181 (–3052 to 114)	0.30 (0.16 to 0.43)	dominant	5	—	95	—
CUA	–1181 (–3052 to 114)	0.0183 (.0182 to 0.0185)	dominant	5	—	95	—
Intervention costs minus 50%							
CEA	–1271 (–3068 to 18)	0.30 (0.16 to 0.43)	dominant	3	—	97	—
CUA	–1271 (–3068 to 18)	0.0183 (.0182 to 0.0185)	dominant	3	—	97	—
Completer analysis							
CEA	–1169 (–2963 to 625)	0.32 (0.18 to 0.46)	dominant	9	—	91	—
CUA	–1169 (–2963 to 625)	0.0188 (0.0089 to 0.028)	dominant	9	—	91	—
Public health care perspective							
Intervention costs plus 20%							
CEA	246 (–59 to 405)	0.30 (0.16 to 0.43)	831 (59 – 1778)	98	—	2	—
CUA	246 (–59 to 405)	0.0183 (0.0182 to 0.0185)	14,380 (1135 to 31,826)	98	—	2	—
Intervention costs plus 50%							
CEA	336 (38 to 503)	0.30 (0.16 to 0.43)	1129 (306 to 2137)	100	—	—	—
CUA	336 (38 to 503)	0.0183 (0.0182 to 0.0185)	19,360 (4671 to 40,673)	100	—	—	—
Intervention costs minus 20%							
CEA	127 (–157 to 295)	0.30 (0.16 to 0.43)	413 (–400 to 1185)	87	—	13	—
CUA	127 (–157 to 295)	0.0183 (0.0182 to 0.0185)	7826 (–4120 to 22,565)	90	—	10	—
Intervention costs minus 50%							
CEA	37 (–258 to 201)	0.30 (0.16 to 0.43)	129 (–636 to 838)	66	—	34	—
CUA	37 (–258 to 201)	0.0183 (0.0182 to 0.0185)	2384 (–9114 to 15,017)	66	—	34	—
Completer analysis							
CEA	206 (–145 to 395)	0.32 (0.18 to 0.46)	656 (–296 to 1725)	92	—	8	—
CUA	206 (–145 to 395)	0.0188 (0.0089 to 0.028)	12,046 (–3920 to 33,174)	92	—	8	—

[a]95% CIs in this column were bias-corrected and accelerated.

XSL•FO
RenderX

[b]ICER: incremental cost-effectiveness ratio.

[c]NEQ: north–east quadrant.

[d]NWQ: north–west quadrant.

[e]SEQ: south–east quadrant.

[f]SWQ: south–west quadrant.

[g]Costs due to presenteeism based on the Health and Labour Questionnaire method.

[h]CEA: cost-effectiveness analysis.

[i]dominant: The intervention resulted in higher effects at lower costs compared to the control.

[j]CUA: cost–utility analysis: quality-adjusted life years.

Discussion

Principal Results

Our study was set out to evaluate the cost-effectiveness and cost-utility of a guided iCBT-I intervention as an adjunct to usual care to reduce insomnia symptoms in schoolteachers in comparison with a wait list control condition with unrestricted access to TAU from a societal and a public health care perspective. Statistically significant differences favoring the intervention group were found for both health outcomes (symptom-free status and QALYs). From a societal perspective, the iCBT-I intervention dominated the wait list-control condition, meaning that the iCBT-I intervention has better health effects for less costs than does usual care in schoolteachers with insomnia. From a public health care perspective, the ICERs were €650 and €11,285 for a symptom-free individual and QALY gained, respectively. At a WTP threshold of €20,000 per QALY gained, the intervention's probability of being cost-effective was 89%.

Comparison With Prior Work

Although the effectiveness of iCBT-I is well established [15], there is a critical gap in health-economic evidence for iCBT-I. To our knowledge, this is the first trial-based economic evaluation of an iCBT-I intervention to reduce insomnia symptoms using a societal and public health care perspective. As such, results from our trial add to the converging evidence pointing to the cost-effectiveness of CBT interventions for insomnia. Thiart et al [20] used the same study to evaluate the cost and benefits of the iCBT-I intervention as seen from an employer's perspective. Results of the current health-economic evaluation line up agreeably with these findings. Baka et al [21] compared a guided iCBT-I intervention to care as usual for insomnia patients in general practice. Analogous to our findings, mean societal costs were lower in the intervention group than in the care as usual group, and, in contrast to our results, the cost–utility analyses revealed a lower probability (69%) of the intervention being cost-effective compared to care as usual at a ceiling ratio of €30,000 per QALY gained. This difference could be due to different types of control conditions used (care as usual vs wait list control). Applying an employer's or societal perspective seems to generate incremental costs favoring interventions groups when participants are employees, or at least in the productive age groups. In addition, our findings match with available health-economic evidence from a recent systematic review (N=7) showing that CBT-I was cost-effective compared to pharmacotherapy or no treatment [19].

Our findings from the public health care perspective showing that the iCBT intervention resulted in better health effects but achieved this at higher costs are also in line with findings from this systematic review [19]. Three trial-based economic evaluations employing a public health care perspective showed that CBT-I led to greater health improvements at higher costs compared to either TAU [59,60] or a wait list control condition [61], with time horizons ranging from 8 weeks to 6 months. In terms of QALYs gained, studies reported a low (34%) [61] and high (99%) [59] probability of CBT-I being cost-effective at a maximum WTP of £30,000 (€31,727) in the United Kingdom. Watanabe et al [60] reported a 90% chance of CBT-I being cost-effective at a WTP threshold of US$40,000 (€29,400) per QALY gained. In contrast, one study conducted from a public health care perspective showed that CBT-I was cheaper and more effective than TAU. However, there was large uncertainty around cost estimates resulting in a moderate probability (70%) of being cost-effective at a WTP of £30,000 (€31,727) [62].

Limitations

This study has some limitations. First, the time horizon of this study was limited to 6 months. However, results of an economic model investigating the long-term cost-effectiveness of CBT-I among long-term hypnotic drug users with chronic sleep difficulties compared to TAU indicated that any increase in the timeframe of the economic evaluation produces substantial reductions in the incremental costs per QALY gained. The cost-effectiveness of CBT-I improved even when treatment effects were reduced radically over time [59]. Further studies should thus assess the long-term clinical and cost-effectiveness of iCBT-I to reaffirm its long-term cost-effectiveness. Second, the iCBT-I intervention was compared to a wait list control condition in the present trial. Although patients in the control group had full access to treatment as usual, we cannot rule out a potential nocebo effect in the wait list control condition [63]. In addition, pharmacoeconomic guidelines recommend standard care (eg, face-to-face CBT-I) as comparator [64]. Future studies should thus directly compare the cost-effectiveness of iCBT-I versus face-to-face CBT-I. Third, although the sample size in this trial was sufficient to demonstrate clinical effectiveness, it needs emphasizing that much larger sample sizes are required for hypothesis testing in economic studies due to the skewness of costs relative to normally distributed health effects [65]. Therefore, it is recommended that future studies employ larger sample sizes to allow for better evaluation of cost changes and sustainability of interventions like iCBT-I. Fourth, we only used the SF-6D to compute utilities and QALYs. However, the choice for the SF-6D (rather than its alternative the EQ-5D-3L [66]) matters, even to the point where decision-makers have to regard

XSL•FO
RenderX

a new intervention as cost-effective or not [67]. In the current study, we flanked the cost–utility analysis (with QALYs) by a cost-effectiveness analysis (with improvements in insomnia) and both economic evaluations led to the same conclusion of the iCBT-I intervention being the preferred option. Nevertheless, future studies should employ different instruments to compute QALYs. Fifth, our results may only be generalized to professions with similar characteristics, such as flexible working hours, loose boundaries between work and private life, and work-home interference. Sixth, the trial has been conducted in a highly educated sample. Hence, we cannot predict the uptake of such an intervention in participants with a lower education level or among those with lower-income status. A recent individual participant data meta-analysis revealed, however, that education was not associated with differential treatment effects of an iCBT intervention to prevent depression [68] although other evidence suggests that better treatment adherence is predicted by higher education [69]. Attrition has been suggested to be an issue, especially in internet-delivered interventions [70]. It is thus warranted to conduct research into the willingness of specific population segments to fully engage in such interventions (ie, how uptake and adherence rates of iCBT interventions could be increased among individuals with less education). Finally, the research context might have led to a self-selection of individuals who might have been more motivated and committed to engage in the iCBT-I intervention than is assumed outside a research context. As a result, findings might not be generalizable to the wider population, but might be representative of specifically those willing to use internet-based interventions in the first place.

Clinical Implications and Future Research

American and European guidelines recommend CBT as the first-line treatment for insomnia disorder [10,11]. Our study supports this recommendation by showing that a guided iCBT-I intervention may reduce insomnia symptoms and improve health-related quality of life. iCBT interventions for mental disorders have often been introduced as potential cost-saving alternatives to face-to-face individual or group therapy [71,72]. Findings from our study add to the evidence base that delivering cognitive-behavioral therapy over the internet has a high probability of being cost-effective in reducing insomnia symptoms among employees. In view of scarce resources and rising costs in health care systems, evidence-based guidance regarding cost-effectiveness of iCBT-I can potentially help to inform decision-makers to the choice of first-line treatments of insomnia. However, future studies should directly compare iCBT-I with face-to-face–delivered CBT-I.

Considering the shortage of therapists and available resources [13], a rather low uptake of professional help [3], and the potential to scale up iCBT-I interventions to efficiently alleviate the health and economic burden caused by insomnia, it would be worthwhile to integrate this type of intervention into routine practice. However, some risks need to be taken into account when scaling up this intervention. Involving individual eCoaches may hamper scaling up the intervention. The support offered by an eCoach may not only affect clinical effectiveness and cost-effectiveness of the intervention but also the target group's willingness to participate in this intervention, thereby influencing intervention effects at the population level. Thus, future studies should compare the acceptability, effectiveness, and cost-effectiveness of guided and unguided iCBT-I interventions. In addition, there are no guarantees that adherence and (by proxy) effectiveness will be maintained if this sort of unsupported iCBT-I intervention is scaled up in the population. Finally, the same technical resources available in the research setting (eg, stable and secure internet connections) may not be available when the intervention is scaled up.

Conclusions

Findings from our trial indicate that iCBT could be considered as good value for money in insomnia therapy. Given the evidence for the effectiveness of iCBT-I interventions to reduce insomnia symptoms, the potential scalability and cost-effectiveness of these interventions might strategically pave the way to alleviate the health and economic burden related to insomnia disorder (and its sequelae) in an affordable way. However, before a nationwide dissemination can be considered, future studies need to evaluate the comparative clinical and economic outcomes of guided and unguided iCBT-I interventions, and to determine what works best for whom such that the deployment of the intervention can optimally target the right population segments. Moreover, implementation studies are needed to clarify the real-world effects of these interventions and to gain insights into the willingness of specific population segments to fully engage in them (eg, individuals with a low education level or those of low-income status).

Acknowledgments

DDE, DL, and MB obtained the funding for this study. DDE, DL, HR, MB, and HH contributed to the design of the clinical study. HH and DDE analyzed the clinical data. CB designed and performed the health-economic evaluation with support from FS. CB had full access to the clinical and economic data and takes full responsibility for the integrity of the data and the accuracy of the health-economic data analyses. CB drafted the manuscript; FS, KS, DL, and DDE supervised the writing process. All authors contributed to the further writing of the manuscript and approved the final manuscript.

This was not an industry-supported study. The European Union funded this study (European Regional Development Fund: ZW6-80119999 and CCI 2007DE161PR001). The funder did not have a role in study design; data collection, analysis, and interpretation of results; or the decision to publish the study results.

XSL•FO
RenderX

Conflicts of Interest

DDE, DL, and MB are stakeholders of the GET.ON Institute, which aims to implement scientific findings related to digital health interventions into routine care. HH serves as business developer at the GET.ON Institute. DDE has served as a consultant on the scientific advisory boards of Sanofi, Novartis, Minddistrict, Lantern, Schoen Kliniken, Ideamed, as well as German health insurance companies (BARMER and Techniker Krankenkasse) and a number of federal chambers for psychotherapy. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Additional figures.
[PDF File (Adobe PDF File), 432 KB - jmir_v23i5e25609_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH Checklist (V.1.6.1).
[PDF File (Adobe PDF File), 11399 KB - jmir_v23i5e25609_app2.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.
2. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
3. Morin CM, LeBlanc M, Bélanger L, Ivers H, Mérette C, Savard J. Prevalence of insomnia and its treatment in Canada. Can J Psychiatry 2011 Sep;56(9):540-548. [doi: 10.1177/070674371105600905] [Medline: 21959029]
4. Wardle-Pinkston S, Slavish DC, Taylor DJ. Insomnia and cognitive performance: A systematic review and meta-analysis. Sleep Med Rev 2019 Dec;48:101205. [doi: 10.1016/j.smrv.2019.07.008] [Medline: 31522135]
5. Kyle SD, Morgan K, Espie CA. Insomnia and health-related quality of life. Sleep Med Rev 2010 Feb;14(1):69-82. [doi: 10.1016/j.smrv.2009.07.004] [Medline: 19962922]
6. Hertenstein E, Feige B, Gmeiner T, Kienzler C, Spiegelhalder K, Johann A, et al. Insomnia as a predictor of mental disorders: a systematic review and meta-analysis. Sleep Med Rev 2019 Feb;43:96-105. [doi: 10.1016/j.smrv.2018.10.006] [Medline: 30537570]
7. Espie CA, Pawlecki B, Waterfield D, Fitton K, Radocchia M, Luik AI. Insomnia symptoms and their association with workplace productivity: cross-sectional and pre-post intervention analyses from a large multinational manufacturing company. Sleep Health 2018 Jun;4(3):307-312. [doi: 10.1016/j.sleh.2018.03.003] [Medline: 29776626]
8. Hillman D, Mitchell S, Streatfeild J, Burns C, Bruck D, Pezzullo L. The economic cost of inadequate sleep. Sleep 2018 Aug 01;41(8):1-13. [doi: 10.1093/sleep/zsy083] [Medline: 29868785]
9. Léger D, Bayon V. Societal costs of insomnia. Sleep Med Rev 2010 Dec;14(6):379-389. [doi: 10.1016/j.smrv.2010.01.003] [Medline: 20359916]
10. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):125-133. [doi: 10.7326/M15-2175] [Medline: 27136449]
11. Riemann D, Baglioni C, Bassetti C, Bjorvatn B, Dolenc Groselj L, Ellis JG, et al. European guideline for the diagnosis and treatment of insomnia. J Sleep Res 2017 Dec;26(6):675-700 [FREE Full text] [doi: 10.1111/jsr.12594] [Medline: 28875581]
12. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
13. Luik AI, van der Zweerde T, van Straten A, Lancee J. Digital delivery of cognitive behavioral therapy for insomnia. Curr Psychiatry Rep 2019 Jun 04;21(7):50 [FREE Full text] [doi: 10.1007/s11920-019-1041-0] [Medline: 31161406]
14. Espie CA. "Stepped care": a health technology solution for delivering cognitive behavioral therapy as a first line insomnia treatment. Sleep 2009 Dec;32(12):1549-1558 [FREE Full text] [doi: 10.1093/sleep/32.12.1549] [Medline: 20041590]
15. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
16. Mohit B, Cohen JT. Trends of cost-effectiveness studies in sleep medicine. Sleep Med 2019 Jan;53:176-180. [doi: 10.1016/j.sleep.2018.06.001] [Medline: 30282589]
17. Botteman M. Health economics of insomnia therapy: implications for policy. Sleep Med 2009 Sep;10 Suppl 1:S22-S25. [doi: 10.1016/j.sleep.2009.07.001] [Medline: 19647482]
18. Wickwire EM, Shaya FT, Scharf SM. Health economics of insomnia treatments: The return on investment for a good night's sleep. Sleep Med Rev 2016 Dec;30:72-82. [doi: 10.1016/j.smrv.2015.11.004] [Medline: 26874067]

XSL•FO
RenderX

19. Natsky AN, Vakulin A, Chai-Coetzer CL, Lack L, McEvoy RD, Lovato N, et al. Economic evaluation of cognitive behavioural therapy for insomnia (CBT-I) for improving health outcomes in adult populations: a systematic review. Sleep Med Rev 2020 Jul 09;54:101351. [doi: 10.1016/j.smrv.2020.101351] [Medline: 32739824]
20. Thiart H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. Sleep 2016 Oct 01;39(10):1769-1778 [FREE Full text] [doi: 10.5665/sleep.6152] [Medline: 27450686]
21. Baka A, Van der Zweerde T, Lancee J, Bosmans J, Van Straten A. PMH31 cost-effectiveness of guided internet-delivered cognitive behavioral therapy in comparison with care-as-usual for patients with insomnia in general practice. Value in Health 2019 Nov;22:S686. [doi: 10.1016/j.jval.2019.09.1505]
22. Thiart H, Lehr D, Ebert DD, Sieland B, Berking M, Riper H. Log in and breathe out: efficacy and cost-effectiveness of an online sleep training for teachers affected by work-related strain--study protocol for a randomized controlled trial. Trials 2013 Jun 11;14:169 [FREE Full text] [doi: 10.1186/1745-6215-14-169] [Medline: 23759035]
23. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015 Mar;41(2):164-174. [doi: 10.5271/sjweh.3478] [Medline: 25590336]
24. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health 2013;16(2):231-250 [FREE Full text] [doi: 10.1016/j.jval.2013.02.002] [Medline: 23538175]
25. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
26. Siebern AT, Manber R. Insomnia and its effective non-pharmacologic treatment. Med Clin North Am 2010 May;94(3):581-591. [doi: 10.1016/j.mcna.2010.02.005] [Medline: 20451034]
27. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007 Apr;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001] [Medline: 17184887]
28. Wells A. Metacognitive Therapy for Anxiety and Depression. New York: The Guilford Press; 2009.
29. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003 Feb;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377] [Medline: 12585811]
30. Kreiner GE, Hollensbe EC, Sheep ML. Balancing borders and bridges: negotiating the work-home interface via boundary work tactics. AMJ 2009 Aug;52(4):704-730. [doi: 10.5465/amj.2009.43669916]
31. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
32. Thorndike FP, Ritterband LM, Saylor DK, Magee JC, Gonder-Frederick LA, Morin CM. Validation of the insomnia severity index as a web-based measure. Behav Sleep Med 2011;9(4):216-223. [doi: 10.1080/15402002.2011.606766] [Medline: 22003975]
33. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004 Sep;42(9):851-859. [doi: 10.1097/01.mlr.0000135827.18610.0d] [Medline: 15319610]
34. Brazier J, Rowen D, Hanmer J. Revised SF-6D scoring programmes: a summary of improvements. PRO newsletter 2008;40:14-15. [doi: 10.4135/9781412971980.n298]
35. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004 Sep;13(9):873-884. [doi: 10.1002/hec.866] [Medline: 15362179]
36. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess 2014 May;18(34):vii-viii, xiii [FREE Full text] [doi: 10.3310/hta18340] [Medline: 24857402]
37. Bouwmans C, De JK, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan SS, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
38. Hakkaart-van RLV, Donker M, Tiemens B. Manual Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P). Rotterdam: Erasmus University; 2002.
39. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 04;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
40. Kählke F, Buntrock C, Smit F, Berking M, Lehr D, Heber E, et al. Economic evaluation of an internet-based stress management intervention alongside a randomized controlled trial. JMIR Ment Health 2019 May 15;6(5):e10866 [FREE Full text] [doi: 10.2196/10866] [Medline: 31094355]

XSL•FO
RenderX

41. Nobis S, Ebert DD, Lehr D, Smit F, Buntrock C, Berking M, et al. Web-based intervention for depressive symptoms in adults with types 1 and 2 diabetes mellitus: a health economic evaluation. Br J Psychiatry 2018 Dec;212(4):199-206. [doi: 10.1192/bjp.2018.10] [Medline: 30071909]
42. Exchange rates (indicator). OECD Data. 2015. URL: https://data.oecd.org/conversion/exchange-rates.htm [accessed 2016-08-30]
43. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics 2008;26(9):733-744. [doi: 10.2165/00019053-200826090-00004] [Medline: 18767894]
44. Krauth C, Hessel F, Hansmeier T, Wasem J, Seitz R, Schweikert B. [Empirical standard costs for health economic evaluation in Germany -- a proposal by the working group methods in health economic evaluation]. Gesundheitswesen 2005 Oct;67(10):736-746. [doi: 10.1055/s-2005-858698] [Medline: 16235143]
45. Bock J, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. Gesundheitswesen 2015 Jan;77(1):53-61. [doi: 10.1055/s-0034-1374621] [Medline: 25025287]
46. Rote Liste 2013. Frankfurt/Main: Rote Liste Service GmbH; 2013:3939192708.
47. Drummond M, Sculpher M, Torrance G, O'Brian B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford: Oxford University Press; 2005.
48. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. Pharmacoeconomics 1992 Jul;2(1):67-76. [doi: 10.2165/00019053-199202010-00008] [Medline: 10146980]
49. Krol M, Brouwer W. Unpaid work in health economic evaluations. Soc Sci Med 2015 Nov;144:127-137. [doi: 10.1016/j.socscimed.2015.09.008] [Medline: 26421997]
50. Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. BMJ 1999 Sep 04;319(7210):635-638 [FREE Full text] [doi: 10.1136/bmj.319.7210.635] [Medline: 10473486]
51. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
52. StataCorp. Stata Statistical Software: Release 13. College Station. Texas: StataCorp LP; 2013.
53. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005 May;14(5):487-496. [doi: 10.1002/hec.944] [Medline: 15497198]
54. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? Journal of Health Economics 2001 Jul;20(4):461-494. [doi: 10.1016/s0167-6296(01)00086-8]
55. Willan AR, Lin DY, Manca A. Regression methods for cost-effectiveness analysis with censored data. Stat Med 2005 Jan 15;24(1):131-145. [doi: 10.1002/sim.1794] [Medline: 15515137]
56. Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press; 2014.
57. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ 2004 May;13(5):405-415. [doi: 10.1002/hec.903] [Medline: 15127421]
58. Drummond M, McGuire A. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001.
59. Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M. Psychological treatment for insomnia in the regulation of long-term hypnotic drug use. Health Technol Assess 2004 Feb;8(8):iii-iv, 1 [FREE Full text] [doi: 10.3310/hta8080] [Medline: 14960254]
60. Watanabe N, Furukawa TA, Shimodera S, Katsuki F, Fujita H, Sasaki M, et al. Cost-effectiveness of cognitive behavioral therapy for insomnia comorbid with depression: analysis of a randomized controlled trial. Psychiatry Clin Neurosci 2015 Jun;69(6):335-343 [FREE Full text] [doi: 10.1111/pcn.12237] [Medline: 25205008]
61. Bonin E, Beecham J, Swift N, Raikundalia S, Brown JS. Psycho-educational CBT-Insomnia workshops in the community. A cost-effectiveness analysis alongside a randomised controlled trial. Behav Res Ther 2014 Apr;55:40-47. [doi: 10.1016/j.brat.2014.01.005] [Medline: 24607500]
62. Tsiachristas A, Waite F, Freeman D, Luengo-Fernandez R. Cost-effectiveness of cognitive-behavioural therapy for sleep disorder added to usual care in patients with schizophrenia: the BEST study. BJPsych Open 2018 May;4(3):126-135 [FREE Full text] [doi: 10.1192/bjo.2018.2] [Medline: 29971156]
63. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014 Sep;130(3):181-192. [doi: 10.1111/acps.12275] [Medline: 24697518]
64. Sacristán JA, Abellán-Perpiñán JM, Dilla T, Soto J, Oliva J. Some reflections on the use of inappropriate comparators in CEA. Cost Eff Resour Alloc 2020 Aug 27;18(1):29 [FREE Full text] [doi: 10.1186/s12962-020-00226-8] [Medline: 32874138]
65. Briggs A. Economic evaluation and clinical trials: size matters. BMJ 2000 Dec 02;321(7273):1362-1363 [FREE Full text] [doi: 10.1136/bmj.321.7273.1362] [Medline: 11099268]
66. Brooks R. EuroQol: the current state of play. Health Policy 1996 Jul;37(1):53-72. [doi: 10.1016/0168-8510(96)00822-6] [Medline: 10158943]

XSL•FO
RenderX

67. Davis JC, Liu-Ambrose T, Khan KM, Robertson MC, Marra CA. SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. Osteoporos Int 2012 Jul;23(7):1849-1857 [FREE Full text] [doi: 10.1007/s00198-011-1770-3] [Medline: 21909728]
68. Reins JA, Buntrock C, Zimmermann J, Grund S, Harrer M, Lehr D, et al. Efficacy and moderators of internet-based interventions in adults with subthreshold depression: an individual participant data meta-analysis of randomized controlled trials. Psychother Psychosom 2020 Jun 16:1-13 [FREE Full text] [doi: 10.1159/000507819] [Medline: 32544912]
69. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
70. 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]
71. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
72. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 01;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]

Abbreviations

(i)CBT-I: (internet-based) cognitive behavioral therapy for insomnia
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CONSORT: Consolidated Standards of Reporting Trials
HLQ: Health and Labor Questionnaire
ICER: incremental cost effect ratio
ISI: Insomnia Severity Index
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
QALY: quality-adjusted life years
SF-6D: 6D Health State Short Form
SFS: symptom-free status
TAU: treatment-as-usual
TiC-P: Treatment Inventory of Costs in Patients with psychiatric disorders
WTP: willingness to pay

Edited by G Eysenbach; submitted 09.11.20; peer-reviewed by J Li; comments to author 30.11.20; revised version received 12.12.20; accepted 17.04.21; published 24.05.21.

Please cite as:
Buntrock C, Lehr D, Smit F, Horvath H, Berking M, Spiegelhalder K, Riper H, Ebert DD
Guided Internet-Based Cognitive Behavioral Therapy for Insomnia: Health-Economic Evaluation From the Societal and Public Health Care Perspective Alongside a Randomized Controlled Trial
J Med Internet Res 2021;23(5):e25609
URL: https://www.jmir.org/2021/5/e25609
doi:10.2196/25609
PMID:34028361

©Claudia Buntrock, Dirk Lehr, Filip Smit, Hanne Horvath, Matthias Berking, Kai Spiegelhalder, Heleen Riper, David Daniel Ebert. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.05.2021. 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

Medical Specialty Recommendations by an Artificial Intelligence Chatbot on a Smartphone: Development and Deployment

Hyeonhoon Lee[1], MS; Jaehyun Kang[2]; Jonghyeon Yeo[3]

[1]Department of Clinical Korean Medicine, Graduate School, Kyung Hee University, Seoul, Republic of Korea

[2]Department of Computer Science, Yonsei University, Seoul, Republic of Korea

[3]School of Computer Science and Engineering, Pusan National University, Busan, Republic of Korea

Corresponding Author:
Hyeonhoon Lee, MS
Department of Clinical Korean Medicine
Graduate School
Kyung Hee University
23 Kyungheedae-ro, Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 29589207
Email: jackli0373@gmail.com

Abstract

Background: The COVID-19 pandemic has limited daily activities and even contact between patients and primary care providers. This makes it more difficult to provide adequate primary care services, which include connecting patients to an appropriate medical specialist. A smartphone-compatible artificial intelligence (AI) chatbot that classifies patients' symptoms and recommends the appropriate medical specialty could provide a valuable solution.

Objective: In order to establish a contactless method of recommending the appropriate medical specialty, this study aimed to construct a deep learning–based natural language processing (NLP) pipeline and to develop an AI chatbot that can be used on a smartphone.

Methods: We collected 118,008 sentences containing information on symptoms with labels (medical specialty), conducted data cleansing, and finally constructed a pipeline of 51,134 sentences for this study. Several deep learning models, including 4 different long short-term memory (LSTM) models with or without attention and with or without a pretrained FastText embedding layer, as well as bidirectional encoder representations from transformers for NLP, were trained and validated using a randomly selected test data set. The performance of the models was evaluated on the basis of the precision, recall, F_1-score, and area under the receiver operating characteristic curve (AUC). An AI chatbot was also designed to make it easy for patients to use this specialty recommendation system. We used an open-source framework called "Alpha" to develop our AI chatbot. This takes the form of a web-based app with a frontend chat interface capable of conversing in text and a backend cloud-based server application to handle data collection, process the data with a deep learning model, and offer the medical specialty recommendation in a responsive web that is compatible with both desktops and smartphones.

Results: The bidirectional encoder representations from transformers model yielded the best performance, with an AUC of 0.964 and F_1-score of 0.768, followed by LSTM model with embedding vectors, with an AUC of 0.965 and F_1-score of 0.739. Considering the limitations of computing resources and the wide availability of smartphones, the LSTM model with embedding vectors trained on our data set was adopted for our AI chatbot service. We also deployed an Alpha version of the AI chatbot to be executed on both desktops and smartphones.

Conclusions: With the increasing need for telemedicine during the current COVID-19 pandemic, an AI chatbot with a deep learning–based NLP model that can recommend a medical specialty to patients through their smartphones would be exceedingly useful. This chatbot allows patients to identify the proper medical specialist in a rapid and contactless manner, based on their symptoms, thus potentially supporting both patients and primary care providers.

(J Med Internet Res 2021;23(5):e27460) doi:10.2196/27460

XSL•FO
RenderX

KEYWORDS

artificial intelligence; chatbot; COVID-19; deep learning; deployment; development; machine learning; medical specialty; natural language processing; recommendation; smartphone

Introduction

The COVID-19 pandemic has encouraged the development of telemedicine and the use of digital platforms [1]. In the field of remote medical support, various digital tools help minimize the number of face-to-face interactions between patients and health care providers (HCPs) [2]. Artificial intelligence (AI) chatbots, also called conversational agents, have recently been designed to support HCPs [3]. Most AI chatbots utilize deep learning–based natural language processing (NLP), which can analyze natural human language input and respond appropriately in a conversational manner [4]. The advantages of an AI chatbot over human HCPs include the absence of face-to-face interaction; minimization of bias based on certain patient demographic characteristics such as age, gender, and race; greater cost-effectiveness; and 24/7 availability since the chatbot does not get fatigued or sick [5]. In a recent systematic review on the effectiveness of AI chatbots in health care, the bots performed well in terms of both usability and satisfaction, and overall positive or mixed effectiveness was reported in most studies [6].

In primary care, it may be important for HCPs to determine which medical specialty is most appropriate for their patients. Since patients generally have no professional medical knowledge, they have to rely on the decisions made by the primary care provider. It cannot be overemphasized that high-quality primary care systems ensure favorable health outcomes and decrease the economic burden [7]. Currently, however, the COVID-19 pandemic limits the amount of physical contact between patients and primary care providers. This state of affairs hinders communication between patients and primary care providers, which prevents timely provision of appropriate treatment by a medical specialist and worsens health outcomes. Therefore, the need for digital tools including AI chatbots to complement the care provided by HCPs and support the decision-making capacity of primary care providers (for example, by connecting patients to medical specialists) is greater.

To develop AI chatbots, electronic medical records (EMRs) have generally been used as input data pipelines for NLP-related medical studies. Using EMRs for NLP facilitates the identification of patients with digestive disorders [8,9] and the prediction of the risk of psychiatric problems, such as actual self-harm, harm to others or victimization, and the risk of health care–associated infections such as surgical site infections [10,11]. Furthermore, EMRs have been used to develop an excellent medical specialty classifier built using deep learning–based NLP, which reportedly had area under receiver operating characteristic curve (AUC) scores of 0.975 and 0.991 and F_1-scores of 0.845 and 0.870 in 2 different EMR data sets [12]. However, the text-formatted EMR data generated by HCPs are mainly composed of medical terminology, which differs from the expressions used by patients to describe their symptoms. Therefore, a new data set consisting of sentences that are commonly used by patients to ask their HCPs about their symptoms was required to fulfill the purpose of our chatbot. This study aims to collect data that accurately describes patients' symptoms in a real-world setting (much more friendly to patients than HCPs) and to develop a deep learning–based NLP model for medical specialty classification. Specifically, we constructed various deep learning–based NLP models, compared their performances, and then selected the best model for our AI chatbot. Finally, the developed AI chatbot was deployed in Google Cloud, which can be used on both desktops and smartphones.

Methods

Data Collection and Cleansing

For supervised learning of the deep learning–based model for NLP, both a single-sentence symptom description and its corresponding medical specialty were required. A Korean website called HiDoc [13]—a web-based health care platform—provides a medical consultation service for anonymous users (patients) by linking them to more than 4000 medical specialists. All medical specialists submitted their professional licenses for approval to provide medical consultation to HiDoc users. HiDoc posts, in which users describe their symptoms, have two parts: title and content. The titles of the posts, in a single-sentence format, were collected for our data set. The medical specialty corresponding to each title sentence was obtained from the profile of the medical specialist who responded to the post.

In the first step of the data cleansing process, duplicate and missing data were eliminated. Second, ambiguous sentences that were not sufficient for accurate classification of the medical specialty, including sentences with ≤2 words or those not related to medical consultations, were manually excluded. Third, very few instances of mislabeled data were appropriately relabeled by a well-trained physician.

Exploratory Data Analysis

Exploratory data analysis (EDA) was performed to extract the interpretable features of the data before development of the deep learning–based NLP models. First, we enumerated the sentences related to symptoms in each class to assess the data distribution. We also visualized the most frequently used words to create lists of words, such as stop words (not useful for classification) and keywords (useful for classification) for word representation. The sentence lengths were determined (in terms of both word count and character count) to ascertain the maximum length of the input sequence for each model.

Development of Deep Learning Models

Long Short-Term Memory Models

Clinical word representation has proven to be an important factor in the performance of NLP models [14]. To extract the

XSL•FO
RenderX

appropriate word representation, each sentence was mined for nouns excluding the words in the list of stop words, and each noun was converted to an index if included in the list of keywords. Using a tokenizer in the Keras library, 15,000 high-frequency nouns were replaced with the corresponding numbers. Thereafter, padding tokens were added to the sentence to ensure consistency in sentence length. As the input of the long short-term memory (LSTM) models, we used either word embedding vectors trained on our data set or word embedding vectors pretrained with the Korean corpus from FastText. The number of embedding dimensions was set to 2048. As the fundamental LSTM architecture, a 256-cell bidirectional LSTM served as the backbone of the model with or without an added attention layer [15,16]. Further, 2 fully connected layers with a rectified linear unit function were applied, followed by a dense layer with the softmax function for classification.

We built 4 different LSTM models. The first is the LSTM model with embedding vectors trained on our data set. Second is the LSTM model with Bahdanau attention. Most settings are the same in the first and second models, but the second model includes an attention layer with 256 cells after the bidirectional LSTM layer. The third model is the LSTM model with FastText pretrained vectors. We loaded FastText's pretrained vector data set and reorganized it to create the embedding matrix. This matrix was used as the embedding layer's weight. The unmentioned hyperparameters are the same as those in the first LSTM model. The last variation is the LSTM model with both FastText vectors and Bahdanau attention. It has an embedding layer based on FastText pretrained vectors and a Bahdanau attention layer.

When compiling all 4 models, categorical cross-entropy and Adam were applied as the loss function and training optimizer, respectively. For the training process, 10-fold cross-validation was conducted to ensure evaluation accuracy. Furthermore, early stopping callback monitoring validation loss was used to prevent overfitting. The batch size and number of epochs were set to 1000 and 30, respectively.

Bidirectional Encoder Representations From Transformers Model

The bidirectional encoder representations from transformers (BERT) model, proposed by Google, has achieved state-of-the-art performance on biomedical and clinical entity normalization with EMRs as well as other NLP tasks such as question answering and natural language inference [17]. Therefore, we used the BERT model for sentence classification through fine-tuning. For preprocessing of the sentences, we applied an open-source tokenizer from Huggingface [18]. This tokenizer encoded each sentence to be used as input of the BERT model by, for example, adding special tokens ([CLS] for beginning and [SEP] for end of sentence), padding to the maximum sequence length, and generating an attention mask. The BERT classification model was built from the pretrained BERT model with a fully connected layer and the softmax function on top. Categorical cross-entropy and Adam optimizer were used for compilation. As in the LSTM models, 10-fold cross-validation was used to increase the reliability of the statistical findings, and early stopping callback observing "validation loss" with patience 2 was set to avoid overfitting. Training was carried out for 30 epochs per fold, with a batch size of 100.

Evaluation

We performed 10-fold cross-validation for the entire data set for each model, and then we calculated the mean score of the 10 different folds to evaluate the general performance of the models. Model performance was evaluated on the basis of precision, recall, F_1-score, and AUC. These indexes are calculated from the rates of true positive, false positive, and false negative results, as follows:







AUC = area under the curve of the false positive rate (x-axis) vs the true positive rate (y-axis).

The Wilcoxon signed rank test was used to test the significance of between-group differences.

Development of the AI Chatbot

We next developed a user-friendly application that allows patients to interact with our chatbot model. As we aimed to make our chatbot usable for all, without creating any digital health disparities on the basis of factors including age, and ensure that it provided accurate medical information, we adopted the formal tone of the Korean language. A well-designed chatbot provides agility for developers and is able to run continuously in any environment. We constructed a chatbot architecture considering those factors.

A typical chatbot architecture can be simplified into 2 parts. The first part is the client-side showing the main user interface. The other part is the server-side that includes the dialog processing logic and an NLP model.

We developed a prototype chatbot client using an open-source chatbot framework Alpha [19]. There are various chatbot user interface alternatives to Alpha, including "chat-bubble," but Alpha is superior for several reasons. In general, open-source chatbot user interface frameworks only have features for sending and receiving messages. They are difficult to test or execute in a developer environment. The Alpha chatbot framework is a highly customizable, fully complete chatbot framework. Alpha is predockerized and built with a WebKit, which may help developers quickly run and test the continuously changing codebase. Furthermore, Alpha includes a cross-platform feature that allows it to be used on both desktops and smartphones. For rapid development, we modified Alpha's client-side dialog logic to fit with the targeted user base.

Tools

Frequently used words were visualized in a word cloud using a Python package called "WordCloud." The Python package for Korean NLP "KoNLPy" was used for word representation.

XSL•FO
RenderX

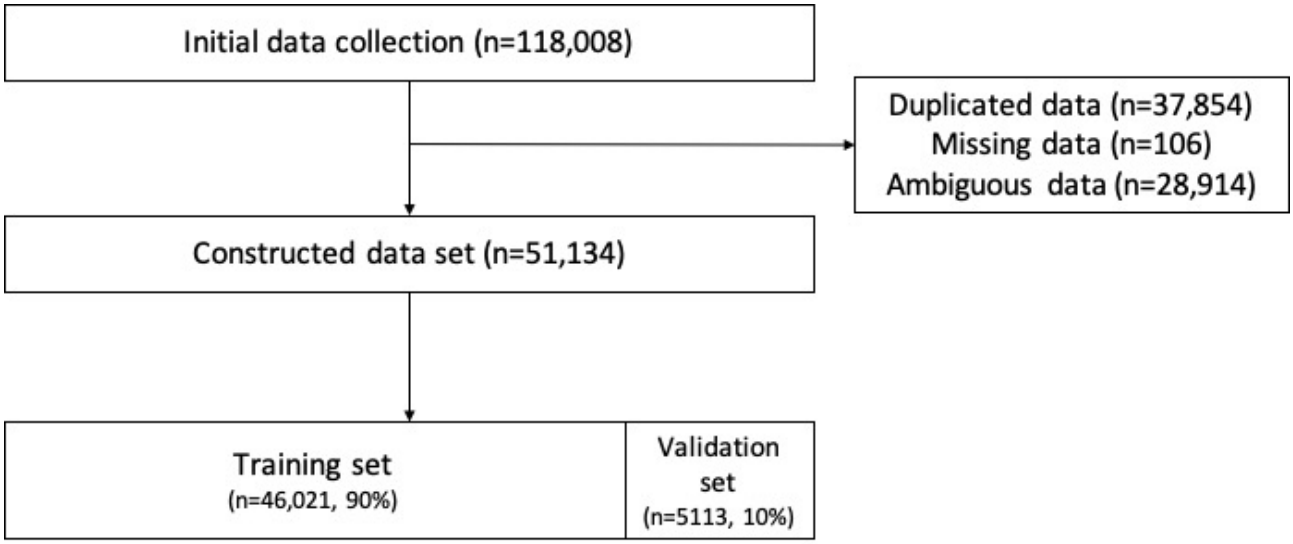
The Huggingface “Transformers” package was used to encode sentences and load a pretrained model for BERT. The “Tensorflow” framework was adopted for building and evaluating the deep learning models. Google Colab, a cloud service for machine learning research, was used in this study. It provides various libraries and frameworks for deep learning and a robust graphics processing unit. Statistical analysis was conducted using R (version 4.0.3, The R Foundation).

Results

Data Set Construction

We initially collected 118,008 sentences that discussed patients’ symptoms, which fit into the 26 classes of medical specialty in HiDoc, and eliminated duplicate data (n=37,854) and missing data (n=106). After excluding ambiguous sentences (n=28,914), the final data set including 51,134 sentences in 26 medical specialty classes was constructed. The data set was randomly split into a training set (n=66,021, 90%) and validation set (n=5,113, 10%) during 10-fold cross-validation. The flow diagram of the entire data set construction process is shown in Figure 1.

Figure 1. Flow diagram for data set construction.



EDA

By specialty, the number of sentences was the highest in dermatology (19.87%), followed by psychology (12.04%), neurology (9.93%), and orthopedic surgery (7.10%) (Table 1).

As the word cloud showed that some frequent words including daily expressions were predominant in the data set, we investigated the list of frequent words. Furthermore, some useful statistical information could be identified at the word and character levels (Table 2).

The EDA results provided 3 crucial ideas to help identify the best NLP model for the AI chatbot. First, owing to the class imbalance of the data, the F_1-score should be considered the most important measure to accurately evaluate and compare the models. Second, to improve the performance of the models through data preprocessing and tokenizing, we decided to compile a list of stop words, such as “hello,” “question,” and “ask,” which might not be helpful for classification, and a list of keywords, which are frequently found in medical expressions (n=15,000), such as “pain,” “head,” and “sudden.” Third, considering the sentence length at both levels, the maximum length of the input sequences for each deep learning model (10 for the LSTM models and 30 for the BERT model) was determined to fix the shape of the input layer in each model.

XSL•FO
RenderX

Table 1. Number of sentences describing symptoms in 26 medical specialty classes.

Medical specialty	Sentences, n (%)
Dermatology	10172 (19.89)
Psychology	6154 (12.04)
Neurology	5080 (9.93)
Orthopedic surgery	3628 (7.10)
Gastroenterology	3096 (6.05)
Otorhinolaryngology	3065 (5.99)
Ophthalmology	2852 (5.58)
Neurosurgery	2028 (3.97)
Rehabilitation medicine	1934 (3.78)
Cardiology	1640 (3.21)
Pulmonology	1334 (2.61)
Plastic surgery	1292 (2.53)
Korean traditional medicine	1227 (2.40)
Obstetrics and gynecology	1170 (2.29)
Infectious disease	1115 (2.18)
Dentistry	1100 (2.15)
Endocrinology	970 (1.90)
Cardiothoracic surgery	713 (1.39)
Rheumatology	654 (1.28)
Urology	521 (1.02)
Anesthesiology	418 (0.82)
Nephrology	283 (0.55)
Hematology and oncology	273 (0.53)
Allergy and immunology	227 (0.44)
General surgery	117 (0.23)
Emergency medicine	71 (0.14)

Table 2. Features of sentence length at the word and character level.

Statistic	Word level, n	Character level, n
Maximum	52	156
Minimum	1	1
Mean	4.68	20.14
Median	4	18
SD	2.78	11.03
First quartile	3	12
Third quartile	6	27

Comparison of Deep Learning Models

The performance outcomes after 10-fold cross-validation in the 5 different deep learning models for NLP are summarized in Table 3. The BERT model showed the best performance, followed by the LSTM model, with embedding vectors trained on our data set.

After saving all trained models in the server, we determined that the BERT model was too heavy to be run on our GCP computation engine with limited performance because the server should be able to handle requests on both desktops and smartphones. Therefore, we had to utilize the lighter LSTM model with embedding vectors trained on our data set, which showed the second-best classification performance.

XSL•FO
RenderX

Table 3. Comparison of classification performance among the 5 different deep learning models during 10-fold cross-validation.

Model#	Word embedding	Model	Precision (95% CI)	Recall (95% CI)	F_1-score (95% CI)	Area under the receiver operating characteristic curve (95% CI)	*P* value
1	Trained on our own data set	Long short-term memory	0.805 (0.800-0.810)	0.686 (0.684-0.689)	0.739 (0.737-0.742)	0.965 (0.964-0.966)	<.01
2	Trained on our own data set	Long short-term memory+attention	0.798 (0.794-0.801)	0.672 (0.668-0.675)	0.727 (0.725-0.730)	0.959 (0.957-0.960)	<.01
3	Pretrained from FastText	Long short-term memory	0.789 (0.786-0.791)	0.622 (0.617-0.627)	0.693 (0.689-0.696)	0.963 (0.962-0.964)	Reference[a]
4	Pretrained from FastText	Long short-term memory+attention	0.800 (0.796-0.803)	0.645 (0.638-0.651)	0.711 (0.707-0.716)	0.965 (0.964-0.966)	<.01
5	Pretrained from Transformers	Bidirectional encoder representations from transformers	0.799 (0.795-0.803)	0.740 (0.737-0.743)	0768 (0.766-0.769)	0.964 (0.963-0.965)	<.01

[a]Comparison of F_1-scores of the models with the reference, which showed the lowest F_1-score.

Deployed AI Chatbot

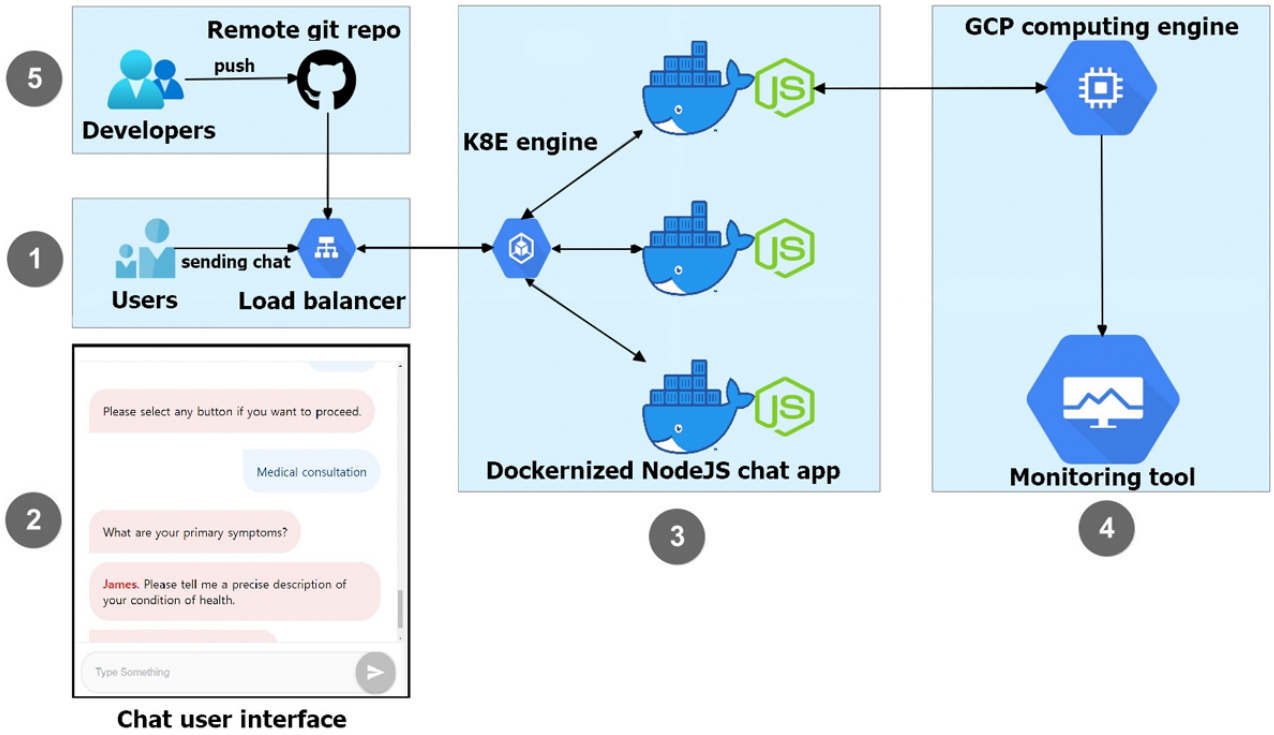
Figure 2 presents the overall architecture of our chatbot. On top of the complete chatbot features, we added basic bubble-like buttons that classify medical departments using our deep learning model, link to an online appointment system, list medical doctors, and provide a simple introduction of the chatbot. When a user selects “medical consultation,” the chatbot asks to send a natural Korean language sentence that describes the user's health state. The sentence is sent to a deployed server and the NLP model classifies medical specialties using the given information. The server responds to the client with the classified medical specialty. The user's device shows the complete sentence generated on the client-side. Based on the medical specialty outcome, the chatbot asks the user if it should provide some services, such as presenting the schedule of physicians in the medical specialty and helping to make an appointment with a physician.

Our NLP model was deployed to a server-side e2-medium (2 vCPU, 4 GB memory) GCP computation engine. We deployed this chatbot framework on Google Cloud Platform Kubernetes Engine (GKE), which is conveniently coupled with containerized by docker, for rapid development. GKE's Kubernetes has a load-balancer feature that distributes application traffic to prevent sudden chatbot outages. Applying Kubernetes at an early stage of development helps developers optimize the uptime, performance, and cost of the application.

We also applied automated git automation with Kubernetes to accelerate the speed of delivery. When a developer pushes modified code from a local computer to a remote GitHub repository, GitHub triggers the containerized chat app cloned in GKE. This continuous deployment method eases the burden on developers by eliminating the need for complex configured scripts on a deployment server. In the cloud-native environment, this chatbot framework laid the groundwork to easily extend further features.

XSL•FO
RenderX

Figure 2. Architecture of the chatbot. This figure illustrates the workflow of the developed prototype chatbot. (1) Users send a sentence through the chatbot's input box. (2) A simple example of our chatbot's user interface. (3) The Containerized Node js chat application includes responding logic without classifying a medical department. (4) The GCP computing engine, which has a natural language processing model, which classifies the medical department from the sentence input in (3). (5) This dockerized NodeJS-based chat app is deployed by continuous git push steps.



Discussion

Principal Findings

In this study, the BERT model was the best classifier of medical specialty, followed by the LSTM model with embedding vectors trained on our data set. These 2 models had an AUC of 0.964 and 0.965 and F_1-score of 0.768 and 0.739, respectively; however, these values were lesser than those reported previously (AUC of 0.975-0.991 and F_1-score of 0.845-0.870) [12]. Nonetheless, the main findings of this study include the following: (1) we developed not only a deep learning classification model for medical sentences but also a prototype AI chatbot to be executed on a smartphone; (2) to our knowledge, this study is the first to use real patients' actual descriptions of their symptoms to develop a deep learning–based NLP model for medical specialty classification; and (3) we constructed Korean language data sets, which can be used in further studies.

This AI chatbot may help patients understand which medical specialty is appropriate for the treatment of their present symptoms and then make an appointment with the corresponding medical specialist, without any face-to-face contact throughout the process. Previous deep learning–based studies on medical sentences developed and suggested optimized deep learning models for several purposes [12,20-23]. However, even with these deep learning models, it takes a great deal of time, effort, and trial and error to deploy a web-based service for practical use. We also made the final choice of deep learning model considering performance and size before deployment. Similar to various recently developed web-based medical tools that offer assessment, communication, management, and other features [24-28], we also focused on the accessibility of our AI chatbot among users with smartphones. As the use of AI techniques in medical services such as telemedicine is growing owing to the COVID-19 pandemic, we believe that this type of research (from the development of a deep learning model to deployment of an AI chatbot) has become extremely important as it can be rapidly applied in practical medical settings.

Most NLP-related medical studies have used EMRs such as clinical notes and discharge notes [21,29,30]. These types of medical texts consist of "physician-friendly" words; that is, medical terminology. Therefore, in those studies, patients' expressions of their own complaints must be transformed in an appropriate manner before being used as input for the model. It was also not suitable for the input of the AI chatbot. However, the data set constructed and used for the development of the deep learning–based NLP model in this study consists of "patient-friendly" words taken from a currently operating web-based health care platform. The use of this data set helped develop an AI chatbot with which patients can interact easily and conveniently. Furthermore, this data set of Korean medical language may be useful for further NLP-based medical studies, including those on diagnosis, treatment, and prediction [31-33] because different languages have different features that can greatly influence the study of NLP.

Limitations

One limitation of this study is that the corpus of the data set that was used for the development of the NLP models was obtained

XSL•FO
RenderX

from one specific website. Demographically, most HiDoc users who accessed the website were less than 65 years of age (Multimedia Appendix 1). They also have nonemergent symptoms that are not complex enough to be differentially diagnosed and not too sensitive to readily share with others. This may explain why the largest proportion of symptom-related sentences were associated with dermatology, followed by psychology (Table 1). Second, the present chatbot, which uses text messaging for remote communication, might be uncomfortable for older adults. To not ignore the needs of this vulnerable and continuously increasing population, an additional communication strategy, such as voice-based communication, is required. A recent study reported that a smart speaker was an effective solution for bringing an AI-based digital health system for older Korean adults [34]. Therefore, other AI-based technologies including smart speakers should be considered for further studies. Third, our model provides classification into 26 medical specialties based on the system of a single general hospital. This may affect the generalizability of our model; a few medical specialties that exist in other medical facilities may be missing. Fourth, some aspects of the study may have been restricted, such as the amount or quality of the data and computing resources. To improve the performance of the deep learning model for classification, additional large collections of high-quality data as well as a costly, latest-generation high-performance computation engine could be used for a large model such as BERT. Fifth, concerning the interpretability of deep learning models, the use of shallow learning models such as support vector machines and naïve Bayes classifiers might be suggested for further studies. Sixth, a clinical trial is required to evaluate how well our AI chatbot service chooses the correct medical specialty in a real-world setting.

Conclusions

In this study, we illustrate the potential of a smartphone-compatible AI chatbot service to recommend a suitable medical specialty to patients. We developed a deep learning–based NLP model and deployed the novel AI chatbot. This type of non–face-to-face medical service is a promising strategy to overcome the current difficulties associated with the COVID-19 pandemic.

Authors' Contributions

HL designed the study. HL, JK, and JY implemented the study. HL drafted the manuscript. JK and JY provided critical revisions to the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic data of HiDoc users.
[DOCX File , 21 KB - jmir_v23i5e27460_app1.docx]

References

1. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
2. Pappot N, Taarnhøj GA, Pappot H. Telemedicine and e-Health Solutions for COVID-19: Patients' Perspective. Telemed J E Health 2020 Jul;26(7):847-849. [doi: 10.1089/tmj.2020.0099] [Medline: 32329654]
3. Powell J. Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test. J Med Internet Res 2019 Oct 28;21(10):e16222 [FREE Full text] [doi: 10.2196/16222] [Medline: 31661083]
4. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
5. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
6. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
7. Shi L. The impact of primary care: a focused review. Scientifica (Cairo) 2012;2012:432892 [FREE Full text] [doi: 10.6064/2012/432892] [Medline: 24278694]
8. Zand A, Sharma A, Stokes Z, Reynolds C, Montilla A, Sauk J, et al. An Exploration Into the Use of a Chatbot for Patients With Inflammatory Bowel Diseases: Retrospective Cohort Study. J Med Internet Res 2020 May 26;22(5):e15589 [FREE Full text] [doi: 10.2196/15589] [Medline: 32452808]
9. Ananthakrishnan AN, Cai T, Savova G, Cheng S, Chen P, Perez RG, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. Inflamm Bowel Dis 2013 Jun;19(7):1411-1420 [FREE Full text] [doi: 10.1097/MIB.0b013e31828133fd] [Medline: 23567779]

XSL•FO
RenderX

10. Le DV, Montgomery J, Kirkby KC, Scanlan J. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. J Biomed Inform 2018 Oct;86:49-58 [FREE Full text] [doi: 10.1016/j.jbi.2018.08.007] [Medline: 30118855]
11. Chandran D, Robbins DA, Chang C, Shetty H, Sanyal J, Downs J, et al. Use of Natural Language Processing to identify Obsessive Compulsive Symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. Sci Rep 2019 Oct 02;9(1):14146 [FREE Full text] [doi: 10.1038/s41598-019-49165-2] [Medline: 31578348]
12. Weng W, Wagholikar KB, McCray AT, Szolovits P, Chueh HC. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. BMC Med Inform Decis Mak 2017 Dec 01;17(1):155 [FREE Full text] [doi: 10.1186/s12911-017-0556-8] [Medline: 29191207]
13. HiDoc. URL: https://www.hidoc.co.kr/ [accessed 2021-04-30]
14. Yetisgen-Yildiz M, Pratt W. The effect of feature representation on MEDLINE document classification. AMIA Annu Symp Proc 2005:849-853 [FREE Full text] [Medline: 16779160]
15. Jagannatha AN, Yu H. Bidirectional RNN for Medical Event Detection in Electronic Health Records. Proc Conf 2016 Jun;2016:473-482 [FREE Full text] [doi: 10.18653/v1/n16-1056] [Medline: 27885364]
16. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 2019 Apr;337:325-338. [doi: 10.1016/j.neucom.2019.01.078]
17. Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT)-Based Models on Large-Scale Electronic Health Record Notes: An Empirical Study. JMIR Med Inform 2019 Sep 12;7(3):e14830 [FREE Full text] [doi: 10.2196/14830] [Medline: 31516126]
18. Transformers. The Hugging Face Team. URL: https://huggingface.co/transformers/ [accessed 2021-01-26]
19. IcaliaLabs/alpha. GitHub. URL: https://github.com/IcaliaLabs/alpha [accessed 2021-01-26]
20. Lin C, Karlson EW, Canhao H, Miller TA, Dligach D, Chen PJ, et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. PLoS One 2013;8(8):e69932 [FREE Full text] [doi: 10.1371/journal.pone.0069932] [Medline: 23976944]
21. McCoy TH, Castro VM, Cagan A, Roberson AM, Kohane IS, Perlis RH. Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study. PLoS One 2015;10(8):e0136341 [FREE Full text] [doi: 10.1371/journal.pone.0136341] [Medline: 26302085]
22. Hughes M, Li I, Kotoulas S, Suzumura T. Medical Text Classification Using Convolutional Neural Networks. Stud Health Technol Inform 2017;235:246-250. [Medline: 28423791]
23. Li Y, Wang X, Hui L, Zou L, Li H, Xu L, et al. Chinese Clinical Named Entity Recognition in Electronic Medical Records: Development of a Lattice Long Short-Term Memory Model With Contextualized Character Representations. JMIR Med Inform 2020 Sep 04;8(9):e19848 [FREE Full text] [doi: 10.2196/19848] [Medline: 32885786]
24. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and Functionalities of Smartphone Apps Related to COVID-19: Systematic Search in App Stores and Content Analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
25. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
26. Lee EW, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, et al. The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study. J Med Internet Res 2020 Jul 07;22(7):e17451 [FREE Full text] [doi: 10.2196/17451] [Medline: 32673252]
27. Philip P, Dupuy L, Morin CM, de Sevin E, Bioulac S, Taillard J, et al. Smartphone-Based Virtual Agents to Help Individuals With Sleep Concerns During COVID-19 Confinement: Feasibility Study. J Med Internet Res 2020 Dec 18;22(12):e24268 [FREE Full text] [doi: 10.2196/24268] [Medline: 33264099]
28. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical Integration of a Smartphone App for Patients With Chronic Pain: Retrospective Analysis of Predictors of Benefits and Patient Engagement Between Clinic Visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
29. Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, et al. Natural language processing of clinical notes for identification of critical limb ischemia. Int J Med Inform 2018 Mar;111:83-89 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.024] [Medline: 29425639]
30. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
31. Heo TS, Kim YS, Choi JM, Jeong YS, Seo SY, Lee JH, et al. Prediction of Stroke Outcome Using Natural Language Processing-Based Machine Learning of Radiology Report of Brain MRI. J Pers Med 2020 Dec 16;10(4):286 [FREE Full text] [doi: 10.3390/jpm10040286] [Medline: 33339385]
32. Jang G, Lee T, Hwang S, Park C, Ahn J, Seo S, et al. PISTON: Predicting drug indications and side effects using topic modeling and natural language processing. J Biomed Inform 2018 Nov;87:96-107 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.015] [Medline: 30268842]

XSL•FO
RenderX

33. Nam Y, Kim H, Kho H. Differential diagnosis of jaw pain using informatics technology. J Oral Rehabil 2018 Aug;45(8):581-588. [doi: 10.1111/joor.12655] [Medline: 29782036]
34. Kim J, Shin E, Han K, Park S, Youn JH, Jin G, et al. Efficacy of Smart Speaker-Based Metamemory Training in Older Adults: Case-Control Cohort Study. J Med Internet Res 2021 Feb 16;23(2):e20177 [FREE Full text] [doi: 10.2196/20177] [Medline: 33591276]

Abbreviations

AI: artificial intelligence
AUC: area under receiver operating characteristic curve
BERT: bidirectional encoder representations from transformers
EDA: exploratory data analysis
EMR: electronic medical record
GKE: Google Cloud Platform Kubernetes Engine
HCP: health care provider
LSTM: long short-term memory
NLP: natural language processing

Edited by C Basch; submitted 26.01.21; peer-reviewed by Z Su, SL Lee; comments to author 01.03.21; revised version received 03.03.21; accepted 17.04.21; published 06.05.21.

Please cite as:
Lee H, Kang J, Yeo J
Medical Specialty Recommendations by an Artificial Intelligence Chatbot on a Smartphone: Development and Deployment
J Med Internet Res 2021;23(5):e27460
URL: https://www.jmir.org/2021/5/e27460
doi:10.2196/27460
PMID:33882012

©Hyeonhoon Lee, Jaehyun Kang, Jonghyeon Yeo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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

Addressing Biodisaster X Threats With Artificial Intelligence and 6G Technologies: Literature Review and Critical Insights

Zhaohui Su[1*], PhD; Dean McDonnell[2*], PhD; Barry L Bentley[3*], PhD; Jiguang He[4], PhD; Feng Shi[5], PhD; Ali Cheshmehzangi[6,7], PhD; Junaid Ahmad[8], PhD; Peng Jia[9,10], PhD

[1]Center on Smart and Connected Health Technologies, Mays Cancer Center, School of Nursing, UT Health San Antonio, San Antonio, TX, United States

[2]Department of Humanities, Institute of Technology Carlow, Carlow, Ireland

[3]Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, United Kingdom

[4]Centre for Wireless Communications, University of Oulu, Oulu, Finland

[5]Department of Research and Development, Shanghai United Imaging Intelligence, Shanghai, China

[6]Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, China

[7]Network for Education and Research on Peace and Sustainability, Hiroshima University, Hiroshima, Japan

[8]Prime Institute of Public Health, Peshawar Medical College, Peshawar, Pakistan

[9]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

[10]International Institute of Spatial Lifecourse Epidemiology, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Peng Jia, PhD
Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
181 Chatham Road South, Kowloon
Hong Kong
China
Phone: 86 2766 5956
Email: jiapengff@hotmail.com

Abstract

Background: With advances in science and technology, biotechnology is becoming more accessible to people of all demographics. These advances inevitably hold the promise to improve personal and population well-being and welfare substantially. It is paradoxical that while greater access to biotechnology on a population level has many advantages, it may also increase the likelihood and frequency of biodisasters due to accidental or malicious use. Similar to “Disease X” (describing unknown naturally emerging pathogenic diseases with a pandemic potential), we term this unknown risk from biotechnologies “Biodisaster X.” To date, no studies have examined the potential role of information technologies in preventing and mitigating Biodisaster X.

Objective: This study aimed to explore (1) what Biodisaster X might entail and (2) solutions that use artificial intelligence (AI) and emerging 6G technologies to help monitor and manage Biodisaster X threats.

Methods: A review of the literature on applying AI and 6G technologies for monitoring and managing biodisasters was conducted on PubMed, using articles published from database inception through to November 16, 2020.

Results: Our findings show that Biodisaster X has the potential to upend lives and livelihoods and destroy economies, essentially posing a looming risk for civilizations worldwide. To shed light on Biodisaster X threats, we detailed effective AI and 6G-enabled strategies, ranging from natural language processing to deep learning–based image analysis to address issues ranging from early Biodisaster X detection (eg, identification of suspicious behaviors), remote design and development of pharmaceuticals (eg, treatment development), and public health interventions (eg, reactive shelter-at-home mandate enforcement), as well as disaster recovery (eg, sentiment analysis of social media posts to shed light on the public’s feelings and readiness for recovery building).

Conclusions: Biodisaster X is a looming but avoidable catastrophe. Considering the potential human and economic consequences Biodisaster X could cause, actions that can effectively monitor and manage Biodisaster X threats must be taken promptly and proactively. Rather than solely depending on overstretched professional attention of health experts and government officials, it is perhaps more cost-effective and practical to deploy technology-based solutions to prevent and control Biodisaster X threats.

XSL•FO
RenderX

This study discusses what Biodisaster X could entail and emphasizes the importance of monitoring and managing Biodisaster X threats by AI techniques and 6G technologies. Future studies could explore how the convergence of AI and 6G systems may further advance the preparedness for high-impact, less likely events beyond Biodisaster X.

(J Med Internet Res 2021;23(5):e26109) doi:10.2196/26109

KEYWORDS

6G; artificial intelligence; biodisaster X; biodisasters; biosafety; biosurveillance; biotechnology; bioterrorism; COVID-19; disease X; sixth-generation technologies

Introduction

Humans have been living with disasters for thousands of years. Records from ancient civilizations, albeit difficult to come across and piecemeal, document the undeniable presence and the lingering impacts of disasters throughout human history [1-8]. Evidence from recent decades alone, for instance, shows that the world witnesses approximately 400 natural disasters and 30-40 armed conflicts annually [9]. Although already daunting, these numbers likely underrepresent the accurate scale and severity of these disasters in society; many often occur in remote places that draw little to no attention of the media or the research community. While humanity is no stranger to various disaster-led debates and discussions focused on the risks posed by natural and anthropogenic disasters [10-12], collectively, research still indicates that societies at large perform poorly when it comes to disaster preparedness [13-19].

As apparent from the COVID-19 pandemic, from the measures assembled during the first outbreaks in 2020 to the faltering waves of public health policies established throughout the first months of 2021, the lack of disaster preparedness and readiness in the wake of global health crises resulted in drastic consequences to the economy and to society [20-25]. Even though the World Health Organization made Disease X (a placeholder term for any unknown pathogenic disease with a pandemic potential) a global priority in 2018 [26], and warned of the potential for a COVID-19–like pandemic to cause disorder in society, unfortunately, few measures were taken to prepare for it [27]. As of May 6, 2021, approximately 155 million COVID-19 cases and over 3.24 million mortalities have been reported worldwide [28]. Parallel with the growth of scientific research in the biological sciences, warnings of the importance of pandemic preparation have been issued throughout the twentieth century [26,29-33] across multiple disciplines. The possibility of bioterrorism and a pathogen's potential to be unleashed without countermeasures is an increasing concern [34-36]. The use of bioweapons by state actors has long been a risk [37-45]; however, with rapid developments in biotechnology, it becomes increasingly feasible for nonstate actors to develop powerful bioweapons in low-barrier contexts, such as home environments, thus increasing the likelihood of biodisasters [46].

Overall, following the pattern of "Disease X," here we introduce the term "Biodisaster X" to refer to disasters caused by the accidental or intentional misuse of biotechnologies by state or nonstate actors. In general, there are no practical insights on approaches to monitor and manage Biodisaster X preemptively. Considering the potential risks Biodisaster X could pose to society's safety and security, research is urgently needed to bridge the knowledge gap. Advances in information and communication technologies have revealed several opportunities to harness digital tools in this effort, particularly with the latest generation of artificial intelligence (AI) techniques and the development of 6G wireless communication technologies, which arguably includes the most promising advanced technological platforms for addressing Biodisaster X threats. Therefore, to this end, this study aimed to identify (1) the potential dangers Biodisaster X poses to society at large and (2) solutions that use emerging 6G and AI technologies to help monitor and manage Biodisaster X threats.

Methods

Methods Overview

A review of the literature on the application of AI and 6G technologies for monitoring and managing biodisasters was conducted via PubMed, using articles published from database inception through to November 16, 2020. The literature search focused on three themes: AI techniques, 6G technologies, and biodisasters. The search terms we utilized are listed in Table 1. To ensure that up-to-date evidence was included, validated news reports, articles identified by examining the reference lists of eligible articles, and records found through an updated literature search in January 2021 were also included.

XSL•FO
RenderX

Table 1. Search terms used on PubMed.

Theme	Search term
Artificial intelligence	"artificial intelligence"[MeSH[a]] OR "artificial intelligence"[TIAB[b]] OR "machine learning"[MeSH] OR "machine learning"[TIAB] OR "deep learning"[MeSH] OR "deep learning"[TIAB]
6G	"sixth-generation communication*"[MeSH] OR "sixth-generation communication*"[TIAB] OR "sixth-generation network*"[MeSH] OR "sixth-generation network*"[TIAB] OR "sixth-generation technolog*"[MeSH] OR "sixth-generation technolog*"[TIAB] OR "sixth-generation cellular" [MeSH] OR "sixth-generation cellular" [TIAB] OR "6G communication*"[MeSH] OR "6G communication*"[TIAB] OR "6G network*"[MeSH] OR "6G network*"[TIAB] OR "6G technolog*"[MeSH] OR "6G technolog*"[TIAB] OR "6G wireless"[MeSH] OR "6G wireless"[TIAB] OR "6G cellular"[MeSH] OR "6G cellular"[TIAB]
Biodisasters	biodisaster*[MeSH] OR biodisaster*[TIAB] OR bio-disaster*[MeSH] OR bio-disaster*[TIAB] OR bioterror*[MeSH] OR bioterror*[TIAB] OR "biological warfare"[MESH] OR "biological warfare"[TIAB] OR "chemical warfare"[MeSH] OR "chemical warfare"[TIAB] OR "bacterial infections and mycoses"[MeSH] OR "bacterial infections and mycoses"[TIAB] OR "virus disease*"[MeSH] OR "virus disease*"[TIAB] OR "parasitic disease*"[MeSH] OR "parasitic disease*"[TIAB] OR "biological threat*"[MeSH] OR "biological threat*"[TIAB] OR bioterror* OR biowar*

[a]MeSH: Medical Subject Headings.

[b]TIAB: limit to title or abstract.

Inclusion and Exclusion Criteria

The literature inclusion criteria for studies in this literature review are listed in Table 2. Overall, articles were excluded if they (1) were not published in English, (2) did not focus on biodisasters, (3) did not focus on either AI techniques or 6G technologies, or (4) did not provide detailed information on the utilization of AI techniques or 6G technologies for monitoring or managing biodisaster threats.

Table 2. Study inclusion criteria.

Data type	Inclusion criteria
Language	English
Study context	Centered on discussions about biodisasters (disasters that occur as a result of infectious pathogens with bioweapon potentials unleashed by state or nonstate actors)
Technology type	Artificial intelligence and 6G
Study design	Focus on utilizing artificial intelligence techniques or 6G technologies to address issues associated with biodisasters

Results

Drawing insights from the review of the literature, the broader literature, as well as the structure of our research objectives and overall study foundation, a clear and comprehensive comparison of the similarities and differences between Disease X and Biodisaster X was developed, details of which are provided in Table 3. As our research objectives were focused on the real-world impact and implications of Disease X and Biodisaster X, rooted in, yet above and beyond what has been discussed and debated in the current literature, our results have been arranged as such. Additionally, to provide a structured and systematic understanding of the findings, we further modeled the study results using the following frames: type of disaster, origin, antecedent, pathogen, transmissibility, transmission predictability, controllability and treatability, nonpharmaceutical mitigation effort, pharmaceutical solution, primary goal, as well as positive unanticipated outcome and negative unintended consequence. A detailed comparison of Disease X and Biodisaster X has been made in Table 3.

XSL•FO
RenderX

Table 3. Similarities and differences between Disease X and Biodisaster X.

Parameter	Disease X	Biodisaster X
Type of Disaster	• Infectious diseases • Natural disasters: epidemics or pandemics	• Infectious diseases • Anthropogenic biodisasters: initially inaccurately identified as a naturally occurring epidemic or pandemic; later identified as anthropogenic
Origin	• None initially, but can be exacerbated by human action or inaction; for example, through livestock mismanagement or failure to contain early infections	• Humans: state or nonstate actors
Antecedent	• None	• Advances in and accessibility of biotechnology; inadequacy of biosecurity measures; intent or malice of nonstate actors
Pathogen	• Originated in nature, with no human engineering • Unfamiliar or unknown to humans	• Originated in a laboratory, mainly as a result of human engineering • Unfamiliar or unknown to humans
Transmissibility	• Highly transmissible, mainly as a result of naturally occurring human interconnectivity	• Highly transmissible, mainly as a result of calculated dissemination and distribution of the pathogen, capitalized on naturally occurring human interconnectivity
Transmission Predictability	• Initially unpredictable; predictability increases over time	• Initially predictable (in principle); extremely low to extremely high predictability over time (depending on the ability of society at large to identify the risk)
Controllability and Treatability	• Low	• Extremely low to extremely high (based on the intent of the actor)
Nonpharmaceutical Mitigation Effort	• Have an agile, evidence-based, and flexible disaster response plan • Equip high-population-density areas with sufficient resources • Limit social interactions; replace physical human interconnectivity with web-based social interactions when possible • Integrate cost-effective technology-based surveillance systems into the emergency management systems • Prepare for a secondary Disease X (potential for mutations or gene transfer, in turn leading to a new Disease X)	• Have an agile, evidence-based, and flexible disaster response plan • Equip high-population density areas with sufficient resources • Limit social interactions; replace physical human interconnectivity with web-based social interactions when possible • Integrate cost-effective technology-based surveillance systems into the emergency management systems • Prepare for a secondary Biodisaster X (actors might further escalate the situation by generating a new Biodisaster X). Once accurately identified as anthropogenic in nature, there will be an urgent need prioritize the identification of the actor
Pharmaceutical Solution	• Difficult and time-consuming to develop vaccines • Extensive and exhaustive efforts needed to identify treatment plans	• Have the potential to be developed fairly easily and time-efficiently once the nonstate actors and the pathogen manufacture process have been identified: vaccine and treatment plans
Primary Goal	• Stop the spread of the pathogen; identify and deploy suitable treatments	• Stop the spread of the pathogen; identify and deploy suitable treatments; locate the source of the pathogen and prevent further action
Positive Unanticipated Outcome and Negative Unintended Consequence	• Physical and psychological health issues associated with the pharmaceutical and nonpharmaceutical solutions • Posttraumatic stress • Increased resource allocation for medicine and public health • Possible economic consequences including a global recession	• Physical and psychological health issues associated with the pharmaceutical and nonpharmaceutical solutions • Posttraumatic stress • Increased resource allocation for medicine, public health, and law enforcement • Possible economic consequences including a global recession

XSL•FO
RenderX

Discussion

Biodisaster X could result in catastrophic human and economic consequences. Yet to date, no studies have examined the potential role of information technologies in preventing and mitigating Biodisaster X, especially in the context of advanced and emerging technologies such as AI and 6G. To bridge the research gap, in this study, we sought to explore 2 fundamental research questions that could considerably enrich the literature: (1) what Biodisaster X might entail and (2) solutions that use AI and emerging 6G technologies to help monitor and manage Biodisaster X threats. In the following sections, we will detail our findings pivoting on these 2 research objectives, as well as practical and powerful strategies that have the potential to effectively control and contain Biodisaster X threats.

Biodisaster X: What is in a Name?

A disaster can be defined as "a serious disruption of the functioning of a community or society involving widespread human, material, economic, or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources" [47]. Based on the contributing causes, disasters are usually categorized as natural (eg, earthquakes, infectious disease-inducing epidemics, or pandemics of natural origin) and anthropogenic (eg, armed conflicts, nuclear accidents, or the release of pathogenic genetically modified organisms from laboratory settings). In the context of this study, biodisasters are defined as disasters that occur as a result of infectious pathogens with bioweapon potential, which are unleashed by state or nonstate actors accidentally and intentionally (eg, the Japanese government's controversial decision to dump Fukushima's contaminated water into the boundless and borderless ocean shared by all life forms on earth, including humans and sharks [48]). In the context of biodisasters, a state actor often takes the form of a nation that deliberately and systematically designs and develops infectious pathogens with its national interest in mind. In contrast, a nonstate actor is an individual or group acting independently to obtain or manufacture a pathogen either owing to misguidance or malice. Of note, although existing multilateral agreements prohibit the production and use of bioweapons by state actors (termed biowarfare) [49], the presence of signed agreements does not imply that accidental or intentional development and release of pathogens by state actors will not occur.

The concept of "bioterrorism," defined as the deliberate release of pathogens that could cause illnesses and deaths in society, is not the focus of this study because "bioterrorism" entails both deliberation and malice (eg, to elicit terror to the public) [50]; antecedents may not necessarily apply to Biodisaster X threats. Insights from behavioral science [51-53] and evidence regarding individual-caused mass casualty events (eg, indiscriminate mass shootings) [54-56] suggest that individual actors' behaviors, potentially leading to the onset of Biodisaster X, may or may not include conscious deliberation to harm. In other words, while it is possible that individual actors' malicious actions might cause some biodisasters, it is also possible that some individual-caused biodisasters are accidental.

Furthermore, the term bioterrorism is limited, in that "terror" is the main outcome. We believe that for Biodisaster X, which could upend lives, livelihoods, and economies, "disaster" is a more appropriate description that sheds light on the scale and severity of its consequences and is more diverse than "terror." Drawing insight from real-world examples, similar to the prevalent ransomware hacks, it is possible that state or individual actors could develop and utilize infectious pathogens as "ransomgens" for financial gain rather than merely aiming to generate terror in society. Therefore, under the current research context, we adopted the term "biodisaster" instead of "bioterrorism." Furthermore, considering that various studies have discussed approaches to address state actor–initiated biodisasters [57-61], this study focuses on biodisasters that are infectious in nature, caused by individual actors, and can result in catastrophic human and economic consequences.

Biodisaster X vs Disease X

The risk of biodisasters, such as Biodisaster X, is increasing in likelihood: advances in technology, particularly the availability and maturity of biotechnology, have grown considerably in recent years. Inadvertently, these advances may resemble those of Oppenheimer [62] in facilitating the release of destructive factors. One example of the misuse of biotechnology is a microbiologist, vaccinologist, and senior biodefense researcher who worked at the United States Army Medical Research Institute of Infectious Diseases, who allegedly engineered the 2001 anthrax attacks [63-65]. While the scale of the 2001 anthrax attacks was minor, it demonstrated how easily biodisasters can occur and how unprepared society was for these events. As seen in the lack of adequate preparation and coherent responses to infectious disease–induced pandemics, including COVID-19 [66-69], Biodisaster X's effects may be compounded to the same, if not greater, degree by incompetence across international, national, and regional agencies and organizations.

The concept of Biodisaster X can be best understood in contrast with Disease X. In terms of similarities, both Biodisaster X and Disease X are driven by pathogens unknown to humans and have the potential to cause crippling effects on society. Furthermore, based on previous inadequacies in response to emergency events including pandemics [66-74], the world at large may be ill-prepared for both Biodisaster X and Disease X. In terms of unique attributes, compared to Disease X, Biodisaster X is more likely to have the following characteristics: (1) having a pathogen directly affiliated to a laboratory; (2) having distinctive and engineered attributes tailored by the capabilities and intentions of the developer; and (3) the origin, development, and history can be definitively ascertained upon identification of the developer, which is not possible for naturally occurring pathogens (eg, the 1918 influenza pandemic), where there is always uncertainty regarding the origin and evolutionary history of the disaster [75-77].

The Imperative of Preparing for Biodisaster X

Some of the deadliest pandemics—the most recent ones ranging from AIDS, severe acute respiratory syndrome, Middle East respiratory syndrome, Ebola, and COVID-19—all have zoonotic origins [78]. Studies have further shown that for viruses that

XSL•FO
RenderX

can transmit from animals to humans, especially those that can infect a diverse range of host species, the transmission speeds are substantially amplified once human-to-human transmission is established, and the diseases can quickly evolve into global pandemics [79]. Consequently, once a pathogen is transmissible within a population, there is a low access threshold: an individual actor can "obtain" these deadly pathogens without the need for advanced laboratory skills or extensive financial resources. However, costs to physical and mental health may reveal a counternarrative.

Based on available evidence, it is difficult to determine whether an individual can be a malicious "patient zero"; an individual who intentionally contracts a novel virus intending to cause infectious disease outbreaks in a society [80]. It is not impossible to purposely study and capture known or unknown deadly pathogens that can trigger infectious diseases; microbial surveys are commonly conducted to identify novel pathogens before they pose a threat to public health [81-84]. In theory, there could be individual actors, with adequate knowledge or experience (similar to the microbiologist allegedly behind the 2011 anthrax attacks [63-65]), who may take the same actions but with different motives, ranging from scientific curiosity to ill-guided intentions. Considering the rich biodiversity of wildlife, along with the large number of "missing viruses" and "missing zoonoses" that remain unidentified [85], close contacts with latent deadly pathogens are nearly impossible to control, which in turn, renders it challenging to locate or identify individual actors who might utilize them. Advances in synthetic biology may further compound the situation, especially considering the scholarly endeavors using pathogens in laboratory settings, which could amount to the level of real-world pandemics (eg, laboratory-cultured viruses such as smallpox [86-88]). The likelihood of Biodisaster X increases in proportion to these factors.

Overall, considering the species diversity of wildlife, the unknown factors related to the scale and severity of viruses in animals, which have the latent potential to infect humans, and the varying degrees of competency of community health centers in detecting infectious disease outbreaks in a bottom-up manner, it could be tremendously difficult for health experts and government officials to monitor potentially emerging Biodisaster X threats. However, not all hope is lost. Technology-based solutions, especially those utilizing AI and 6G technologies, can help address these issues.

The Need for Advanced Technology Solutions for Monitoring and Managing Biodisaster X

The Need for Technology-Based Solutions

Once Biodisaster X becomes a reality, human contact will drive transmission and become the primary fuel for exacerbating infections and deaths caused by the disaster. As seen during the COVID-19 pandemic, owing to virus spread and subsequent public health policies (eg, lockdowns), many critical societal functions could be substantially disrupted. The potential to control and contain human and economic consequences of Biodisaster X, such as the functionality of the health care systems (eg, infected health care professionals) [89-91], may also become critically undermined. In these circumstances, technology-based solutions could be the key to addressing these crises, as they are different from conventional solutions; they are not highly dependent on physical interactions and transportation. Overall, technology-based solutions require limited human resources (eg, with the ability to operate without human input), can be delivered independent of physical human contact (eg, web-based and remote deployment), and are immune to infectious diseases (eg, can function in contaminated environments). Furthermore, technology-based solutions are less vulnerable to issues ranging from physical fatigue to mental health burdens, which are health challenges that frontline workers often face amid emergency events.

The Need for Advanced Technologies

To effectively predict, control, and manage Biodisaster X, which is an event with a low probability (ie, difficult to detect preemptively) and a high impact (ie, difficult to control and contain), advanced technologies are needed. While many emerging technologies can address the dangers and damages associated with Biodisaster X [92,93], 2 families of advanced technology-based solutions show particular promise, namely AI techniques and 6G technologies.

Unique Capabilities of AI

AI is generally considered synonymous with "thinking machines" [94], or techniques that can facilitate "a computer to do things which, when done by people, are said to involve intelligence" [95]. With AI technologies, machines can identify patterns too intricate for humans to identify and process quickly. AI techniques are widely used in areas such as natural language processing, speech recognition, machine vision, targeted marketing, and health care, including efforts to combat COVID-19 [96-99]. While technologies such as virtual reality, smart sensors, drones, and robotics could play a positive role in supporting health care professionals to cope with the pandemic [100-102], AI technologies are arguably most instrumental in addressing some of the most prominent issues health experts and government officials are faced with, ranging from pandemic surveillance to COVID-19 drug and vaccine development [103-106].

AI and machine learning techniques are particularly valuable in their ability to identify trends and patterns across large amounts of data promptly and cost-effectively; for example, in identifying or searching for specific patterns. With natural language processing, for instance, data can be extracted retrospectively from clinical records or prospectively in real time and statistically processed for insights, which, in turn, can supplement existing structured data to enrich actionable information [86]. During the COVID-19 pandemic, natural language processing models have been used to analyze publicly available information such as tweets, tweet timestamps, and geolocation data, to identify and map potential COVID-19 cases cost-effectively, without utilizing testing devices or other medical resources that involve health care professional [107].

Overall, most, if not all, AI techniques are irreplaceable in regard to administering complex tasks such as extracting useful information from large data sets. Moreover, with the

XSL•FO
RenderX

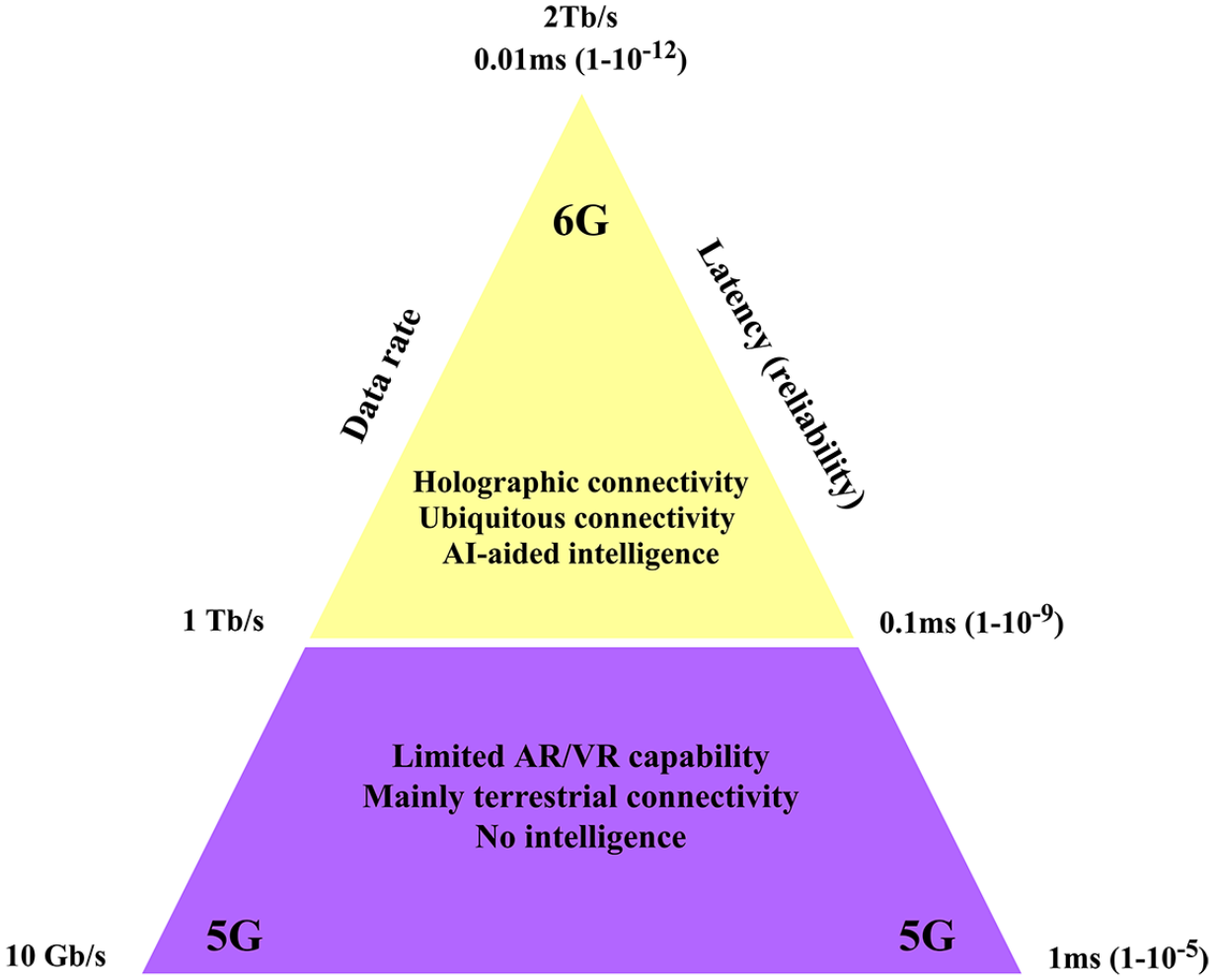
continuously increasing speed of its technological advancements and applications, AI technologies are often utilized as core components in other emerging technologies [108]. Smart sensors that perform advanced tasks, such as effectively identifying and recognizing captured motions and images, often need to integrate deep learning technologies (a subgroup of AI) [109-111]. These combined insights suggest that AI techniques have great potential in monitoring and managing Biodisaster X threats.

Unique Capabilities of 6G Networks

6G technologies are the next generation of wireless communication systems following 5G networks [112]. While 6G is still under development, it is envisioned as the most capable communication network currently available [112-119]. The advantages of 6G networks derive from their high data transmission speed (up to 1 terabyte per second), wireless hyper-connectivity (100 million connections per km^2), low end-to-end latency (< 1 ms), reliability ($1\text{-}10^{-9}$) (reliability in terms of the frame error rate, which is defined as the ratio of the number of incorrectly decoded frames to that of total transmitted frames), and high-accuracy positioning capabilities (indoor: <10 cm in 3D; outdoor: <1 m in 3D) [112-119]. Adding the fact that 6G networks also excel in their energy efficiency and spectrum efficiency, these networks can provide fast and efficient wireless reporting and access to remote computational facilities, facilitating mobile biomonitoring and disaster management.

For instance, the high reliability and data transmission speed of 6G technologies will be of critical importance amid global emergency events with the scale of Biodisaster X. At the onset of the COVID-19 pandemic, many internet companies and service providers experienced outrage and were forced to reduce the amount of data individuals and organizations could utilize to ensure continuous communication for all [120]. This limitation of existing communication networks could compromise the ability of health experts and government officials to monitor and manage COVID-19–related threats and other disasters promptly and properly. Of note, in the face of an extremely deadly, contagious, and fast-developing Biodisaster X, information will be predominantly updated and exchanged remotely and over the internet. The speed and success of updating and exchanging information are highly dependent on the reliability of communication networks, in which 6G technologies excel, especially when spatial big data have been introduced for disease control and prevention since the COVID-19 pandemic [27,108,121]. Figure 1 lists visual comparisons in communication capabilities between 6G and 5G networks.

Figure 1. A schematic representation of the unique advantages of 6G compared to 5G technologies. AI: artificial intelligence, AR: augmented reality, VR: virtual reality.

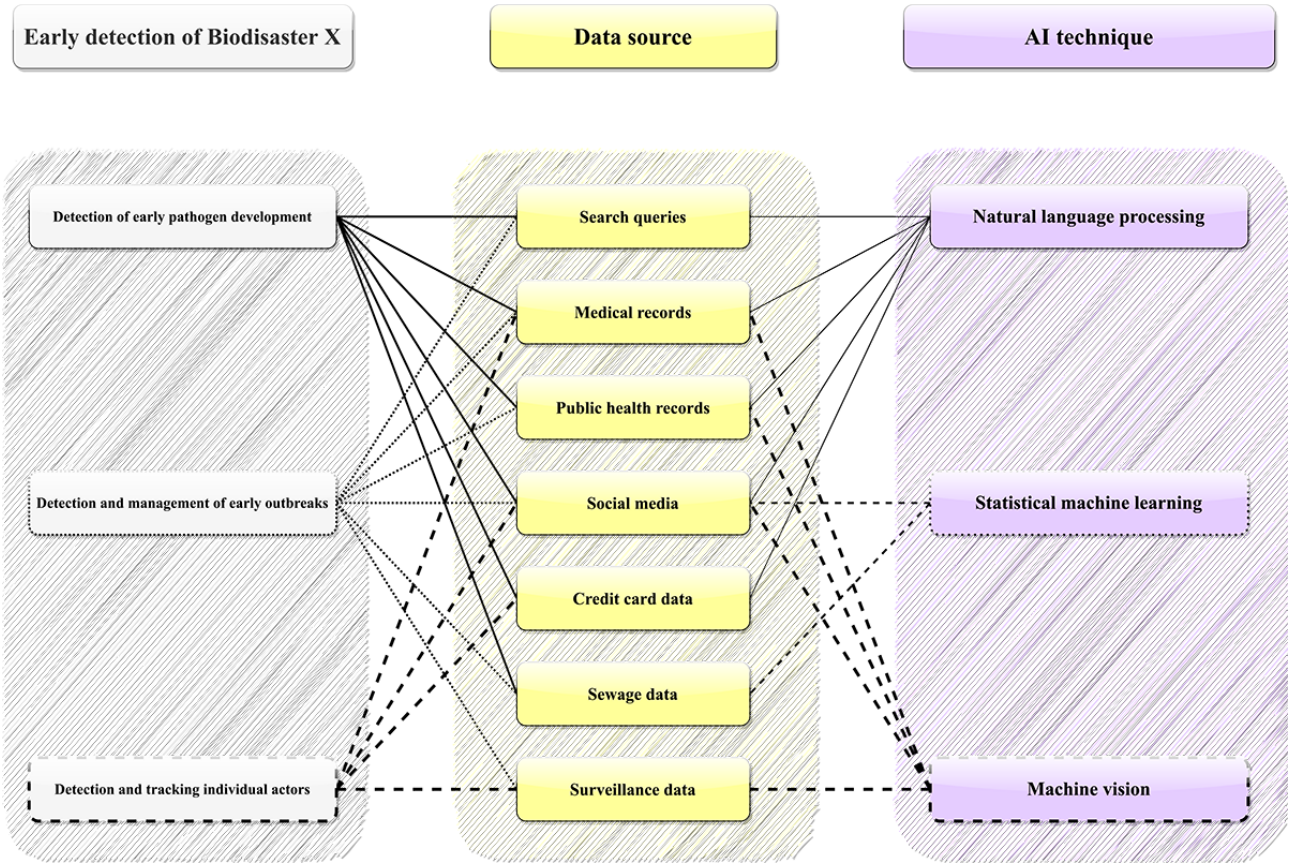


XSL•FO
RenderX

AI and 6G Technologies for Biodisaster X Control and Management

Drawing insights from the COVID-19 pandemic [103-106], AI techniques, especially when coupled with advanced communication capabilities enabled by 6G technologies, can elevate biodisaster control and management. In other words, 6G-based AI technologies can be applied to address issues ranging from early Biodisaster X detection (eg, identification of suspicious behaviors) to remote design and development of pharmaceuticals (eg, treatment development) and public health interventions (eg, reactive shelter-at-home mandate enforcement), and disaster recovery (eg, sentiment analysis of social media posts to shed light on the public sentiments and readiness for recovery building). While there are research issues worth exploring, in this study, we specifically (1) focus on early detection of Biodisaster X, a disaster management stage that could yield maximum benefit in personal and public health protection, and (2) discuss critical aspects of the utilization and application of 6G-based AI technologies in monitoring and managing Biodisaster X threats. Further information on these AI techniques is available in Figure 2.

Figure 2. Artificial Intelligence techniques discussed in detail in the study. Statistical machine learning refers to techniques that do not involve natural language processing or deep learning processes. AI: artificial intelligence.



Natural Language Processing Analysis

With the benefit of high reliability and high data transmission speed, rather than depending on 1 source of data, 6G-based AI surveillance systems could synthesize various data sources to identify suspicious behaviors for epidemiologists, public health experts, and government officials for further analysis; this may also enable the National Intelligent Syndromic Surveillance System (an AI-based disease surveillance system) [108,122]. As available literature suggests [123-126], for instance, with the help of 6G technologies, natural language processing could be applied to analyze search queries (eg, searches such as “How to build an influenza virus weapon”), credit card transaction data (eg, receipts for equipment that are indispensable to develop a deadly virus), medical records (eg, whether the individual actor has a recent or frequent infectious disease diagnosis), public health records (eg, infectious disease cases coinciding with an actor’s activities), and social media activities (eg, reports of infectious disease symptoms and suspicious behaviors) interactively and in real time. While the analysis of isolated individual data sources may yield more noise than useful information, when the analyses are synthesized with corresponding data from diverse and complementary data sources, actionable information could become more readily available.

It is important to note that natural language processing can also help shed light on people’s mental health, such as identifying at-risk populations with severe mental health disorders including schizophrenia and suicide attempts [127-129]. In an analysis of 826,961 unique Reddit users from 2018 to 2020, with the help of natural language processing techniques, researchers have been able to determine how specific mental health disorders (eg, schizophrenia) manifest in textual language, and, in turn, cost-effectively and unobtrusively identified at-risk mental health in social media users [130]. In the same study, natural language processing also offered researchers the opportunity to

XSL•FO
RenderX

compare pre– and mid–COVID-19 posts to identify people whose mental health disorders had become more pronounced during the pandemic [130].

Statistical Machine Learning Solutions

To increase the usefulness of information, 6G-based AI systems can also capitalize on previously tested infectious disease control and management techniques, such as sewage-based epidemiology. Sewage-based epidemiology, or waste-based epidemiology, has helped scientists detect early warnings of pathogenic virus outbreaks, ranging from hepatitis A and norovirus to COVID-19 [131-135]. Different from conventional sewage monitoring, which incorporates little to no innovative technologies, the utilization and application of cost-effective and data-driven technologies in the field have significantly upscaled the information that researchers can gather from wastewater, thus reducing monitoring and reporting times, and, in turn, elevating societies' ability to better monitor and manage infectious diseases [132-135]. The United Kingdom, for instance, has been rapidly developing its sewage monitoring systems to track COVID-19 outbreaks to understand pandemic transmission patterns more swiftly and accurately [136]. By analyzing human waste, researchers in the United Kingdom and worldwide were able to detect RNA from SARS-CoV-2 and better identify the location, scale, and possible trajectory of COVID-19 transmission [136-138].

However, current sewage monitoring technologies often face challenges such as delays in generating insights and inaccurate results owing to limitations caused by poor data processing speed [115,116], which could be overcome by 6G-based AI systems. The use of machine learning techniques has improved the ability to monitor events, such as poliovirus and enterovirus outbreaks with greater precision, as machine learning models could substantially improve surveillance sensitivity and the quantity and quality of actionable information garnered from wastewater [139]. Overall, with the high reliability and high data transmission speed provided by 6G networks, AI systems could advance sewage epidemiology (eg, through widespread remote analysis and evaluation) and facilitate real-time analyses of potential anomalies such as high volume or density of pathogenic viruses in waste and early infectious disease cases, before these threats progress to outbreaks. Insights gained from natural language processing can be further synthesized with information gained through other control measures, such as systematic and comprehensive sewage monitoring (eg, whether there is a suspicious presence of known or suspected pathogens in certain neighborhoods).

Deep Learning–Based Image Analysis

In addition to using natural language processing and statistical machine learning techniques, deep learning–based image analysis can also help with the early detection of Biodisaster X, particularly with the assistance of advanced networks enabled by 6G technologies. One application of deep learning in Biodisaster X monitoring and management would be gauging whether there are patterned disease outbreak signs, on the basis of insights gained from analyzing nontextual health records, such as medical images and real-time surveillance footage. While unstable and low-speed communication networks could affect AI-based disease surveillance architectures that use both regular reporting and remote computation systems, the influences of low-speed networks will be more pronounced for remote systems that use server-side analysis of high-resolution images to generate actionable information. This high network demand may create issues, such as requiring down-sampling of data, in turn compromising discrimination power and accuracy and the ability of the system to yield useful insights [140].

Machine vision enables machines to recognize visuals with the support of cameras, images, and learning models [141]. These processes can be fine-tuned to recognize and identify objects, such as patients' medical images, and help health care professionals better gather information from imaging data and yield diagnoses and treatment plans [142-144]. Combined with the high speed of 6G networks, remote machine vision could be made more effective by allowing the transmission of high-quality images (eg, high resolution) and video-based surveillance (eg, no lag, high frame rate) in real time, facilitating more remote processing than would be possible otherwise.

With 6G technologies, deep learning–based machine vision could have "superhuman" capabilities in monitoring and managing Biodisaster X threats (eg, tracking multiple objects simultaneously and providing real-time analysis of potential threats with high accuracy), helping health officials track and monitor individual actors once identified. With advancements in 6G and deep learning techniques, machine vision may eventually have the potential to identify individual actors, even when they wear face masks. Scientists have proposed the use of machine vision to help health experts and government officials monitor and reinforce COVID-19 safety rules, such as face-making and social distancing (eg, masked face recognition) [145-147].

Limitations

While this study addresses important knowledge gaps in the literature, there are some limitations to consider. First, owing to the scarcity of literature on the subject, we were unable to adopt a systematic review approach to investigate the research question. While our scoping search is justified for the scale of the current studies in this area, it nonetheless limits our findings. Second, while preparation plans for biodiaster X should involve multiple elements, ranging from surveillance, biosecurity mandates, and communication guidelines, to personnel training, we only discussed 1 aspect of Biodisaster X control and prevention. To address these limitations, assuming an increasing number of studies in this field, future studies should adopt a systematic or more comprehensive approach to examine this topic for further understanding ways to better prepare for Biodisaster X threats. Of note, although many AI techniques are now well-established, 6G technologies are still in their infancy. With the many unknown factors associated with AI and 6G development, it is possible that in the future, 6G and AI technologies may underperform in their ability to assist biodisaster monitoring and management, compared to our expectations and discussion in this study. However, this limitation may only be adequately addressed once empirical evidence on real-world 6G-based AI applications becomes more available and accessible.

XSL•FO
RenderX

Conclusions

Biodisaster X is a looming but avoidable catastrophe. Considering the potential human and economic consequences Biodisaster X could have, actions that can effectively monitor and manage Biodisaster X threats must be taken promptly and adequately. Rather than solely depending on overstretched professional attention of health experts and government officials, it is perhaps more cost-effective and practical to deploy technology-based solutions to prevent and control Biodisaster X threats. This study discusses what Biodisaster X could entail and emphasizes the importance of monitoring and management of Biodisaster X threats through AI techniques and 6G technologies. Future studies could investigate how the convergence of AI and 6G systems may further advance the preparedness for high-impact, less likely events beyond Biodisaster X, including the facilitation of the development of the national intelligent syndromic surveillance system.

Acknowledgments

We thank the International Institute of Spatial Lifecourse Epidemiology and the Hong Kong Polytechnic University (1-BE58) for research support, and we wish to express our gratitude to the editors and reviewers for their constructive input.

Authors' Contributions

ZS conceived the study, reviewed the literature, and drafted and edited the manuscript. DMD, BLB, JH, FS, AC, JA, and PJ reviewed the literature and edited the manuscript. All authors approved the final manuscript for submission.

Conflicts of Interest

None declared.

References

1. Tang Q. Ritual Healing: Records of Disasters and ritual function in Li Ji. In: Ritual Civilization and Mythological Coding. Singapore: Springer; Jul 19, 2020:293-319.
2. Touliatis PG. Seismic Disaster Prevention in the History of Structures in Greece. In: Protection of the Architectural Heritage Against Earthquakes. Vienna: Springer; 1996:261-284.
3. Ciottone GR. Introduction to Disaster Medicine. In: Ciottone's Disaster Medicine (2nd Edition). Philadelphia, PA: Elsevier; 2016:2-5.
4. Noji EK. The nature of disaster: general characteristics and public health effects. In: The Public Health Consequences of Disasters. Oxford: Oxford University Press; 1997:3-20.
5. Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, et al. Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. Lancet Infect Dis 2014 Apr;14(4):319-326. [doi: 10.1016/S1473-3099(13)70323-2] [Medline: 24480148]
6. Seger T. The Plague of Justinian and Other Scourges: An analysis of the Anomalies in the Development of the Iron Age population in Finland. Fornvännen 1982;77:184-197 [FREE Full text]
7. Dols MW. Plague in Early Islamic History. J Am Orient Soc 1974 Jul;94(3):371. [doi: 10.2307/600071]
8. Wu T, Perrings C, Kinzig A, Collins JP, Minteer BA, Daszak P. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. Ambio 2017 Feb;46(1):18-29 [FREE Full text] [doi: 10.1007/s13280-016-0809-2] [Medline: 27492678]
9. Waldman R. Natural and human-made disasters. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/eis/field-epi-manual/chapters/Natural-Human-Disasters.html [accessed 2020-08-29]
10. Smith KF, Sax DF, Lafferty KD. Evidence for the role of infectious disease in species extinction and endangerment. Conserv Biol 2006 Oct;20(5):1349-1357. [doi: 10.1111/j.1523-1739.2006.00524.x] [Medline: 17002752]
11. Matheny JG. Reducing the risk of human extinction. Risk Anal 2007 Oct;27(5):1335-1344. [doi: 10.1111/j.1539-6924.2007.00960.x] [Medline: 18076500]
12. Leslie J. The end of the world: The science and ethics of human extinction. London: Routledge; 1996.
13. Alexander DC. Natural Disasters. Boca Raton, FL: CRC Press; Jul 29, 1993.
14. Ritchie H, Roser M. Natural disasters. One World in Data 2019 [FREE Full text]
15. Su Z, McDonnell D, Ahmad J. The need for a disaster readiness mindset: A key lesson from the coronavirus disease 2019 (COVID-19) pandemic. Infect Control Hosp Epidemiol 2021 Jan 25:1-2 [FREE Full text] [doi: 10.1017/ice.2021.26] [Medline: 33487209]
16. Su Z, McDonnell D, Cheshmehzangi A, Li X, Maestro D, Šegalo S, et al. With Great Hopes Come Great Expectations: A Viewpoint on Access and Adoption Issues Associated with COVID-19 Vaccines. JMIR Public Health Surveill 2021 Feb 01 [FREE Full text] [doi: 10.2196/26111] [Medline: 33560997]
17. Su Z, Wen J, Abbas J, McDonnell D, Cheshmehzangi A, Li X, et al. A race for a better understanding of COVID-19 vaccine non-adopters. Brain Behav Immun Health 2020 Dec;9:100159 [FREE Full text] [doi: 10.1016/j.bbih.2020.100159] [Medline: 33052327]

XSL•FO
RenderX

18. Su Z, McDonnell D, Wen J, Kozak M, Abbas J, Šegalo S, et al. Mental health consequences of COVID-19 media coverage: the need for effective crisis communication practices. Global Health 2021 Jan 05;17(1):4 [FREE Full text] [doi: 10.1186/s12992-020-00654-4] [Medline: 33402169]
19. Su Z, Wen J, McDonnell D, Goh E, Li X, Šegalo S, et al. Vaccines are not yet a silver bullet: The imperative of continued communication about the importance of COVID-19 safety measures. Brain Behav Immun Health 2021 Mar;12:100204 [FREE Full text] [doi: 10.1016/j.bbih.2021.100204] [Medline: 33495754]
20. Djalante R, Lassa J, Setiamarga D, Sudjatma A, Indrawan M, Haryanto B, et al. Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. Progress in Disaster Science 2020 Apr;6:100091 [FREE Full text] [doi: 10.1016/j.pdisas.2020.100091]
21. Renda A, Castro R. Towards Stronger EU Governance of Health Threats after the COVID-19 Pandemic. Eur J Risk Regul 2020 Apr 09;11(2):273-282. [doi: 10.1017/err.2020.34]
22. Maani N, Galea S. COVID-19 and Underinvestment in the Public Health Infrastructure of the United States. Milbank Q 2020 Jun;98(2):250-259. [doi: 10.1111/1468-0009.12463] [Medline: 32333418]
23. Ridley EJ, Freeman-Sanderson A, Haines KJ. Surge capacity for critical care specialised allied health professionals in Australia during COVID-19. Aust Crit Care 2021 Mar;34(2):191-193 [FREE Full text] [doi: 10.1016/j.aucc.2020.07.006] [Medline: 32919841]
24. Jia P, Yang S. Are we ready for a new era of high-impact and high-frequency epidemics? Nature 2020 Apr;580(7803):321. [doi: 10.1038/d41586-020-01079-0] [Medline: 32286555]
25. Jia P, Yang S. Time to spatialise epidemiology in China. Lancet Glob Health 2020 Jun;8(6):e764-e765 [FREE Full text] [doi: 10.1016/S2214-109X(20)30120-0] [Medline: 32446343]
26. Prioritizing diseases for research and development in emergency contexts. World Health Organization. URL: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts [accessed 2021-01-18]
27. Jia P. Understanding the Epidemic Course in Order to Improve Epidemic Forecasting. Geohealth 2020 Oct;4(10):e2020GH000303 [FREE Full text] [doi: 10.1029/2020GH000303] [Medline: 33024909]
28. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University: Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2021-05-06]
29. Honigsbaum M. Disease X and other unknowns. Lancet 2019 Apr 13;393(10180):1496-1497 [FREE Full text] [doi: 10.1016/S0140-6736(19)30803-7] [Medline: 30983581]
30. Young PR. Disease X ver1.0: COVID-19. Microbiol Aust 2020;41(2):109. [doi: 10.1071/MA20028]
31. Simpson S, Kaufmann M, Glozman V, Chakrabarti A. Disease X: accelerating the development of medical countermeasures for the next pandemic. Lancet Infect Dis 2020 May;20(5):e108-e115 [FREE Full text] [doi: 10.1016/S1473-3099(20)30123-7] [Medline: 32197097]
32. Osterholm MT. Preparing for the next pandemic. N Engl J Med 2005 May 05;352(18):1839-1842. [doi: 10.1056/NEJMp058068] [Medline: 15872196]
33. Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, et al. Prediction and prevention of the next pandemic zoonosis. Lancet 2012 Dec 01;380(9857):1956-1965 [FREE Full text] [doi: 10.1016/S0140-6736(12)61684-5] [Medline: 23200504]
34. Wagar E. Bioterrorism and the Role of the Clinical Microbiology Laboratory. Clin Microbiol Rev 2016 Jan;29(1):175-189 [FREE Full text] [doi: 10.1128/CMR.00033-15] [Medline: 26656673]
35. Green MS, LeDuc J, Cohen D, Franz DR. Confronting the threat of bioterrorism: realities, challenges, and defensive strategies. Lancet Infect Dis 2019 Jan;19(1):e2-e13 [FREE Full text] [doi: 10.1016/S1473-3099(18)30298-6] [Medline: 30340981]
36. Henderson DA. The looming threat of bioterrorism. Science 1999 Feb 26;283(5406):1279-1282. [doi: 10.1126/science.283.5406.1279] [Medline: 10037590]
37. Nie J, Guo N, Selden M, Kleinman A. Japan's Wartime Medical Atrocities: Comparative Inquiries in Science, History, and Ethics. London: Routledge; 2010.
38. Tsuneishi K. Unit 731 and the human skulls discovered in 1989: Physicians carrying out organized crimes. In: Dark medicine: Rationalizing unethical medical research. Bloomington, IN: Indiana University Press; 2007:73-84.
39. Brody H, Leonard SE, Nie J, Weindling P. U.S. responses to Japanese wartime inhuman experimentation after World War II. Camb Q Healthc Ethics 2014 Apr;23(2):220-230 [FREE Full text] [doi: 10.1017/S0963180113000753] [Medline: 24534743]
40. Nie J. Japanese doctors' experimentation in wartime China. Lancet 2002 Dec;360 Suppl:s5-s6. [doi: 10.1016/s0140-6736(02)11797-1] [Medline: 12504481]
41. Bärnighausen T. Data generated in Japan’s biowarfare experiments on human victims in China, 1932–1945, and the ethics of using them. In: Nie JB, Guo N, Selden M, Kleinman A, editors. Japan's Wartime Medical Atrocities Comparative Inquiries in Science, History, and Ethics. London: Routledge; 2010:81-107.
42. Keiichi T. Unit 731 and the Japanese Imperial Army's Biological Warfare Program. Asia Pac J 2005 Nov 24;3(11):1-9 [FREE Full text]

XSL•FO
RenderX

43. Mellanby K. Medical experiments on human beings in concentration camps in Nazi Germany. Br Med J 1947 Jan 25;1(4490):148-150 [FREE Full text] [doi: 10.1136/bmj.1.4490.148] [Medline: 20244692]
44. Berger RL. Nazi science--the Dachau hypothermia experiments. N Engl J Med 1990 May 17;322(20):1435-1440. [doi: 10.1056/NEJM199005173222006] [Medline: 2184357]
45. Mitscherlich A, Mielke F. Doctors Of Infamy: The Story Of The Nazi Medical Crimes. New York, NY: Normanby Press; 1949.
46. McLeish C, Nightingale P. Biosecurity, bioterrorism and the governance of science: The increasing convergence of science and security policy. Research Policy 2007 Dec;36(10):1635-1654 [FREE Full text] [doi: 10.1016/j.respol.2007.10.003]
47. 2009 UNISDR terminology on disaster risk reduction. United Nations Office for Disaster Risk Reduction. 2009. URL: https://www.undrr.org/publication/2009-unisdr-terminology-disaster-risk-reduction [accessed 2021-05-12]
48. Japan: UN experts 'deeply disappointed' by decision to discharge Fukushima water. UN News. 2021 Apr 15. URL: https://news.un.org/en/story/2021/04/1089852 [accessed 2021-04-26]
49. Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction. United Nations: Office for Disarmament Affairs. URL: http://disarmament.un.org/treaties/t/bwc [accessed 2021-01-20]
50. Jansen HJ, Breeveld FJ, Stijnis C, Grobusch MP. Biological warfare, bioterrorism, and biocrime. Clin Microbiol Infect 2014 Jun;20(6):488-496 [FREE Full text] [doi: 10.1111/1469-0691.12699] [Medline: 24890710]
51. Akerlof GA, Yellen JL. Rational Models of Irrational Behavior. Am Econ Rev 1987;77(2):137-142 [FREE Full text]
52. Mayo E. The Irrational Factor in Human Behavior. Ann Am Acad Political Soc Sci 2016 Sep 08;110(1):117-130 [FREE Full text] [doi: 10.1177/000271622311000112]
53. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
54. Fernandez E, Callen A, Johnson S, Gaspar C, Kulhanek C, Jose-Bueno C. Prevalence, elicitors, and expression of anger in 21st century mass shootings. Aggress Violent Behav 2020 Nov;55:101483 [FREE Full text] [doi: 10.1016/j.avb.2020.101483]
55. Koehler A, Scott RA, Davis R. Surviving the dark night: the Aurora, Colorado, mass shootings. J Emerg Nurs 2014 Sep;40(5):440-445. [doi: 10.1016/j.jen.2014.03.001] [Medline: 24881506]
56. Bjelopera J, Bagalman E, Caldwell S, Finklea K, McCallion G. Public Mass Shootings in the United States: Selected Implications for Federal Public Health and Safety Policy. Congressional Research Service. 2013 Mar 18. URL: https://fas.org/sgp/crs/misc/R43004.pdf [accessed 2021-05-12]
57. Guillemin J. Biological Weapons: From the Invention of State-Sponsored Programs to Contemporary Bioterrorism. New York: Columbia University Press; 2004.
58. Gostin LO, Sapsin JW, Teret SP, Burris S, Mair JS, Hodge JG, et al. The Model State Emergency Health Powers Act: planning for and response to bioterrorism and naturally occurring infectious diseases. JAMA 2002 Aug 07;288(5):622-628. [doi: 10.1001/jama.288.5.622] [Medline: 12150674]
59. Mothershead JL, Tonat K, Koenig KL. Bioterrorism preparedness. III: State and federal programs and response. Emerg Med Clin North Am 2002 May;20(2):477-500. [doi: 10.1016/s0733-8627(02)00004-4] [Medline: 12120488]
60. Koplan J. CDC's strategic plan for bioterrorism preparedness and response. Public Health Rep 2001;116 Suppl 2:9-16 [FREE Full text] [doi: 10.1093/phr/116.S2.9] [Medline: 11880662]
61. Riedel S. Biological warfare and bioterrorism: a historical review. Proc (Bayl Univ Med Cent) 2004 Oct;17(4):400-406 [FREE Full text] [doi: 10.1080/08998280.2004.11928002] [Medline: 16200127]
62. Metropolis N, Rota G, Sharp D, Oppenheimer J. I. Uncommon sense. In: Uncommon Sense. Boston, MA: Birkhäuser Boston; 1984:57-78.
63. Timeline: How The Anthrax Terror Unfolded. NPR. 2011 Feb 15. URL: https://www.npr.org/2011/02/15/93170200/timeline-how-the-anthrax-terror-unfolded?t=1611082987421 [accessed 2021-01-20]
64. Murch RS. "Amerithrax": The Investigation of Bioterrorism Using Bacillus Anthracis Spores in Mailed Letters. Encyclopedia of Bioterrorism Defense 2011 Jul 15:1-6 [FREE Full text] [doi: 10.1002/0471686786.ebd0138]
65. Willman D. The Mirage Man: Bruce Ivins, the Anthrax Attacks, and America's Rush to War. New York, NY: Bantam; Jun 07, 2011.
66. Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in Low- and Middle-Income Countries. JAMA 2020 Apr 28;323(16):1549-1550. [doi: 10.1001/jama.2020.4169] [Medline: 32176764]
67. Truelove S, Abrahim O, Altare C, Lauer SA, Grantz KH, Azman AS, et al. The potential impact of COVID-19 in refugee camps in Bangladesh and beyond: A modeling study. PLoS Med 2020 Jun;17(6):e1003144 [FREE Full text] [doi: 10.1371/journal.pmed.1003144] [Medline: 32544156]
68. Dureab F, Al-Awlaqi S, Jahn A. COVID-19 in Yemen: preparedness measures in a fragile state. Lancet Public Health 2020 Jun;5(6):e311 [FREE Full text] [doi: 10.1016/S2468-2667(20)30101-8] [Medline: 32334648]
69. Senghore M, Savi MK, Gnangnon B, Hanage WP, Okeke IN. Leveraging Africa's preparedness towards the next phase of the COVID-19 pandemic. Lancet Glob Health 2020 Jul;8(7):e884-e885 [FREE Full text] [doi: 10.1016/S2214-109X(20)30234-5] [Medline: 32416768]

XSL•FO
RenderX

70. Garoon J, Duggan P. Discourses of disease, discourses of disadvantage: a critical analysis of National Pandemic Influenza Preparedness Plans. Soc Sci Med 2008 Oct;67(7):1133-1142. [doi: 10.1016/j.socscimed.2008.06.020] [Medline: 18656294]
71. Fineberg HV. Pandemic preparedness and response--lessons from the H1N1 influenza of 2009. N Engl J Med 2014 Apr 03;370(14):1335-1342. [doi: 10.1056/NEJMra1208802] [Medline: 24693893]
72. Su Z, McDonnell D, Ahmad J, Cheshmehzangi A, Li X, Meyer K, et al. Time to stop the use of 'Wuhan virus', 'China virus' or 'Chinese virus' across the scientific community. BMJ Glob Health 2020 Sep;5(9):e003746 [FREE Full text] [doi: 10.1136/bmjgh-2020-003746] [Medline: 32895218]
73. Sánchez-Duque JA, Su Z, Rosselli D, Chica-Ocampo MC, Lotero-Puentes MI, Bolaños-Portilla AM, et al. The ignored pandemic of public health corruption: A call for action amid and beyond SARS-COV-2/COVID-19. JEBAS 2021 Apr 25;9(2):108-116. [doi: 10.18006/2021.9(2).108.116]
74. Su Z, McDonnell D, Li Y. Why is COVID-19 more deadly to nursing home residents? QJM 2021 Jan 16:hcaa343 [FREE Full text] [doi: 10.1093/qjmed/hcaa343] [Medline: 33453109]
75. Oxford JS, Gill D. Unanswered questions about the 1918 influenza pandemic: origin, pathology, and the virus itself. Lancet Infect Dis 2018 Nov;18(11):e348-e354. [doi: 10.1016/S1473-3099(18)30359-1] [Medline: 29935779]
76. Trilla A, Trilla G, Daer C. The 1918 "Spanish flu" in Spain. Clin Infect Dis 2008 Sep 01;47(5):668-673. [doi: 10.1086/590567] [Medline: 18652556]
77. Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. Nat Rev Microbiol 2004 Nov;2(11):909-914 [FREE Full text] [doi: 10.1038/nrmicro1027] [Medline: 15494747]
78. Ye Z, Yuan S, Yuen K, Fung S, Chan C, Jin D. Zoonotic origins of human coronaviruses. Int J Biol Sci 2020;16(10):1686-1697 [FREE Full text] [doi: 10.7150/ijbs.45472] [Medline: 32226286]
79. Kreuder Johnson C, Hitchens PL, Smiley Evans T, Goldstein T, Thomas K, Clements A, et al. Spillover and pandemic properties of zoonotic viruses with high host plasticity. Sci Rep 2015 Oct 07;5:14830 [FREE Full text] [doi: 10.1038/srep14830] [Medline: 26445169]
80. Gill P, Horgan J, Deckert P. Bombing alone: tracing the motivations and antecedent behaviors of lone-actor terrorists,. J Forensic Sci 2014 Mar;59(2):425-435 [FREE Full text] [doi: 10.1111/1556-4029.12312] [Medline: 24313297]
81. Li S. The Compendium of Materia Medica. Beijing: People's Medical Publishing House; 1982.
82. Vongkamjan K, Wiedmann M. Starting from the bench--prevention and control of foodborne and zoonotic diseases. Prev Vet Med 2015 Feb 01;118(2-3):189-195. [doi: 10.1016/j.prevetmed.2014.11.004] [Medline: 25481625]
83. Smith F. On the hunt for the next deadly virus. National Geographic. 2020 Jun 16. URL: https://www.nationalgeographic.com/science/2020/06/coronavirus-on-the-hunt-for-the-next-deadly-virus/ [accessed 2021-01-17]
84. Vanmechelen B, Rector A, Maes P. Virus Hunting: Discovery of New Episomal Circular Viruses by Rolling Circle Techniques. Curr Protoc Microbiol 2017 Feb 06;44:1E.12.1-1E.12.18. [doi: 10.1002/cpmc.23] [Medline: 28166385]
85. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature 2017 Jun 29;546(7660):646-650 [FREE Full text] [doi: 10.1038/nature22975] [Medline: 28636590]
86. Sharma A, Gupta G, Ahmad T, Krishan K, Kaur B. Next generation agents (synthetic agents): Emerging threats and challenges in detection, protection, and decontamination. In: Handbook on Biological Warfare Preparedness. Cambridge, MA: Academic Press; 2020:217-256.
87. Noyce RS, Lederman S, Evans DH. Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. PLoS One 2018;13(1):e0188453 [FREE Full text] [doi: 10.1371/journal.pone.0188453] [Medline: 29351298]
88. Koblentz GD. The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox. Health Secur 2017;15(6):620-628. [doi: 10.1089/hs.2017.0061] [Medline: 28836863]
89. Pietsch B. Central and Southern California have 0 percent I.C.U. capacity, in a state already low on hospital beds. The New York TImes. 2020 Dec 26. URL: https://www.nytimes.com/2020/12/26/world/central-and-southern-california-icu-capacity.html [accessed 2021-01-17]
90. Blumenthal D, Fowler EJ, Abrams M, Collins SR. Covid-19 - Implications for the Health Care System. N Engl J Med 2020 Oct 08;383(15):1483-1488. [doi: 10.1056/NEJMsb2021088] [Medline: 32706956]
91. Cabarkapa S, Nadjidai SE, Murgier J, Ng CH. The psychological impact of COVID-19 and other viral epidemics on frontline healthcare workers and ways to address it: A rapid systematic review. Brain Behav Immun Health 2020 Oct;8:100144 [FREE Full text] [doi: 10.1016/j.bbih.2020.100144] [Medline: 32959031]
92. Jia P. Spatial lifecourse epidemiology. Lancet Planet Health 2019 Feb;3(2):e57-e59 [FREE Full text] [doi: 10.1016/S2542-5196(18)30245-6] [Medline: 30797406]
93. Yang S, Pan X, Zeng P, Jia P. Spatial technologies to strengthen traditional testing for SARS-CoV-2. Trends Microbiol 2021 Mar 14:00062-00067 [FREE Full text] [doi: 10.1016/j.tim.2021.03.003] [Medline: 33775546]
94. Turing A, Braithwaite R, Jefferson G, Newman M. Can Automatic Calculating Machines Be Said To Think? (1952). In: Copeland BJ, editor. The Essential Turing. Oxford: Oxford University Press; 2004:487.
95. Shukla S, Jaiswal V. Applicability of Artificial Intelligence in Different Fields of Life. Int J Sci Eng Res 2013;1(1):28-35 [FREE Full text]
96. Davenport T, Guha A, Grewal D, Bressgott T. How artificial intelligence will change the future of marketing. J Acad Mark Sci 2019 Oct 10;48(1):24-42. [doi: 10.1007/s11747-019-00696-0]

XSL•FO
RenderX

97. Liang B, Yang N, He G, Huang P, Yang Y. Identification of the Facial Features of Patients With Cancer: A Deep Learning-Based Pilot Study. J Med Internet Res 2020 Apr 29;22(4):e17234 [FREE Full text] [doi: 10.2196/17234] [Medline: 32347802]
98. Laghi A. Cautions about radiologic diagnosis of COVID-19 infection driven by artificial intelligence. Lancet Digit Health 2020 May;2(5):e225 [FREE Full text] [doi: 10.1016/S2589-7500(20)30079-0] [Medline: 32373786]
99. Kuziemski M, Misuraca G. AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecomm Policy 2020 Jul;44(6):101976 [FREE Full text] [doi: 10.1016/j.telpol.2020.101976] [Medline: 32313360]
100. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
101. Zeng Z, Chen P, Lew AA. From high-touch to high-tech: COVID-19 drives robotics adoption. Tourism Geographies 2020 May 12;22(3):724-734. [doi: 10.1080/14616688.2020.1762118]
102. Tavakoli M, Carriere J, Torabi A. Robotics, Smart Wearable Technologies, and Autonomous Intelligent Systems for Healthcare During the COVID‐19 Pandemic: An Analysis of the State of the Art and Future Vision. Advanced Intelligent Systems 2020 May 05;2(7):2000071 [FREE Full text] [doi: 10.1002/aisy.202000071]
103. Lin L, Hou Z. Combat COVID-19 with artificial intelligence and big data. J Travel Med 2020 Aug 20;27(5):taaa080 [FREE Full text] [doi: 10.1093/jtm/taaa080] [Medline: 32437541]
104. Bansal A, Padappayil RP, Garg C, Singal A, Gupta M, Klein A. Utility of Artificial Intelligence Amidst the COVID 19 Pandemic: A Review. J Med Syst 2020 Aug 01;44(9):156 [FREE Full text] [doi: 10.1007/s10916-020-01617-3] [Medline: 32740678]
105. Keshavarzi Arshadi A, Webb J, Salem M, Cruz E, Calad-Thomson S, Ghadirian N, et al. Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development. Front Artif Intell 2020;3:65 [FREE Full text] [doi: 10.3389/frai.2020.00065] [Medline: 33733182]
106. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. IEEE Rev Biomed Eng 2021;14:4-15. [doi: 10.1109/RBME.2020.2987975] [Medline: 32305937]
107. Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G. Toward Using Twitter for Tracking COVID-19: A Natural Language Processing Pipeline and Exploratory Data Set. J Med Internet Res 2021 Jan 22;23(1):e25314 [FREE Full text] [doi: 10.2196/25314] [Medline: 33449904]
108. Jia P, Yang S. China needs a national intelligent syndromic surveillance system. Nat Med 2020 Jul;26(7):990. [doi: 10.1038/s41591-020-0921-5] [Medline: 32433591]
109. Guo K, Lu Y, Gao H, Cao R. Artificial Intelligence-Based Semantic Internet of Things in a User-Centric Smart City. Sensors (Basel) 2018 Apr 26;18(5):1341 [FREE Full text] [doi: 10.3390/s18051341] [Medline: 29701679]
110. Guo X, Shen Z, Zhang Y, Wu T. Review on the Application of Artificial Intelligence in Smart Homes. Smart Cities 2019 Aug 02;2(3):402-420. [doi: 10.3390/smartcities2030025]
111. Singh S, Sharma P, Yoon B, Shojafar M, Cho G, Ra I. Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. Sustainable Cities and Society 2020 Dec;63:102364 [FREE Full text] [doi: 10.1016/j.scs.2020.102364]
112. Samsung Research. 6G: The next hyper connected experience for all. everything RF. URL: https://www.everythingrf.com/whitepapers/details/3661-6g-the-next-hyper-connected-experience-for-all [accessed 2021-05-13]
113. Saad W, Bennis M, Chen M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. IEEE Network 2020 May;34(3):134-142. [doi: 10.1109/MNET.001.1900287]
114. Zhou Y, Liu L, Wang L, Hui N, Cui X, Wu J, et al. Service-aware 6G: An intelligent and open network based on the convergence of communication, computing and caching. Digital Communications and Networks 2020 Aug;6(3):253-260 [FREE Full text] [doi: 10.1016/j.dcan.2020.05.003]
115. Nawaz F, Ibrahim J, Awais M, Junaid M, Kousar S, Parveen T. A Review of Vision and Challenges of 6G Technology. IJACSA 2020;11(2). [doi: 10.14569/IJACSA.2020.0110281]
116. Alsharif MH, Kelechi AH, Albreem MA, Chaudhry SA, Zia MS, Kim S. Sixth Generation (6G) Wireless Networks: Vision, Research Activities, Challenges and Potential Solutions. Symmetry 2020 Apr 24;12(4):676. [doi: 10.3390/sym12040676]
117. Dang S, Amin O, Shihada B, Alouini M. What should 6G be? Nat Electron 2020 Jan 24;3(1):20-29. [doi: 10.1038/s41928-019-0355-6]
118. Chowdhury MZ, Shahjalal M, Ahmed S, Jang YM. 6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions. IEEE Open J Commun Soc 2020;1:957-975. [doi: 10.1109/OJCOMS.2020.3010270]
119. Letaief KB, Chen W, Shi Y, Zhang J, Zhang YA. The Roadmap to 6G: AI Empowered Wireless Networks. IEEE Commun Mag 2019 Aug;57(8):84-90. [doi: 10.1109/MCOM.2019.1900271]
120. Kang C, Alba D, Satariano A. Surging traffic is slowing down our internet. The New York Times. URL: https://www.nytimes.com/2020/03/26/business/coronavirus-internet-traffic-speed.html [accessed 2021-01-17]

XSL•FO
RenderX

121. Yang S, Yu C, Jia P. Spatiobehavioral Characteristics - Defining the Epidemiology of New Contagious Diseases at the Earliest Moment Possible. Trends Parasitol 2021 Mar;37(3):179-181 [FREE Full text] [doi: 10.1016/j.pt.2020.12.004] [Medline: 33487571]
122. Jia P, Yang S. Early warning of epidemics: towards a national intelligent syndromic surveillance system (NISSS) in China. BMJ Glob Health 2020 Oct;5(10):e002925 [FREE Full text] [doi: 10.1136/bmjgh-2020-002925] [Medline: 33106238]
123. Alag S. Analysis of COVID-19 clinical trials: A data-driven, ontology-based, and natural language processing approach. PLoS One 2020;15(9):e0239694 [FREE Full text] [doi: 10.1371/journal.pone.0239694] [Medline: 32997699]
124. Ancochea J, Izquierdo JL, Soriano JB. Evidence of Gender Differences in the Diagnosis and Management of Coronavirus Disease 2019 Patients: An Analysis of Electronic Health Records Using Natural Language Processing and Machine Learning. J Womens Health (Larchmt) 2021 Mar;30(3):393-404. [doi: 10.1089/jwh.2020.8721] [Medline: 33416429]
125. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30;23(5):649-685. [doi: 10.1017/S1351324916000383]
126. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 17;349(6245):261-266. [doi: 10.1126/science.aaa8685] [Medline: 26185244]
127. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide Note Classification Using Natural Language Processing: A Content Analysis. Biomed Inform Insights 2010 Aug 04;2010(3):19-28 [FREE Full text] [doi: 10.4137/bii.s4706] [Medline: 21643548]
128. Coppersmith G, Hilland C, Frieder O, Leary R. Scalable mental health analysis in the clinical whitespace via natural language processing. 2017 Presented at: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics; February, 16-19, 2017; Orlando, FL URL: https://ieeexplore.ieee.org/document/7897288/authors#authors [doi: 10.1109/BHI.2017.7897288]
129. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. Comput Math Methods Med 2016;2016:8708434 [FREE Full text] [doi: 10.1155/2016/8708434] [Medline: 27752278]
130. 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]
131. Orive G, Lertxundi U, Barcelo D. Early SARS-CoV-2 outbreak detection by sewage-based epidemiology. Sci Total Environ 2020 Aug 25;732:139298 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.139298] [Medline: 32417555]
132. Ge Z, Song Z, Ding SX, Huang B. Data Mining and Analytics in the Process Industry: The Role of Machine Learning. IEEE Access 2017;5:20590-20616. [doi: 10.1109/ACCESS.2017.2756872]
133. Ansari M, Othman F, Abunama T, El-Shafie A. Analysing the accuracy of machine learning techniques to develop an integrated influent time series model: case study of a sewage treatment plant, Malaysia. Environ Sci Pollut Res Int 2018 Apr;25(12):12139-12149. [doi: 10.1007/s11356-018-1438-z] [Medline: 29455350]
134. Carrascal M, Abian J, Ginebreda A, Barceló D. Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. Sci Total Environ 2020 Dec 10;747:141145. [doi: 10.1016/j.scitotenv.2020.141145] [Medline: 32791406]
135. Guo H, Jeong K, Lim J, Jo J, Kim Y, Park JP, et al. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J Environ Sci (China) 2015 Jun 01;32:90-101. [doi: 10.1016/j.jes.2015.01.007] [Medline: 26040735]
136. Baraniuk C. Sewage monitoring is the UK's next defence against covid-19. BMJ 2020 Jul 15;370:m2599. [doi: 10.1136/bmj.m2599] [Medline: 32669267]
137. Poch M, Garrido-Baserba M, Corominas L, Perelló-Moragues A, Monclús H, Cermerón-Romero M, et al. When the fourth water and digital revolution encountered COVID-19. Sci Total Environ 2020 Nov 20;744:140980 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.140980] [Medline: 32687996]
138. Winn Z. MIT begins testing wastewater to help detect Covid-19 on campus. Massachusetts Institute of Technology. 2020 Oct 02. URL: https://news.mit.edu/2020/testing-wastewater-covid-19-1002 [accessed 2021-02-08]
139. Hamisu AW, Blake IM, Sume G, Braka F, Jimoh A, Dahiru H, et al. Characterizing Environmental Surveillance Sites in Nigeria and Their Sensitivity to Detect Poliovirus and Other Enteroviruses. J Infect Dis 2020 Apr 09:jiaa175. [doi: 10.1093/infdis/jiaa175] [Medline: 32415775]
140. Zhang Q, Zhu S. Visual interpretability for deep learning: a survey. Frontiers Inf Technol Electronic Eng 2018 Jan 28;19(1):27-39. [doi: 10.1631/FITEE.1700808]
141. Forsyth D, Ponce J. Computer vision: A modern approach. Harlow: Pearson Education Limited; 2012.
142. Lemley J, Bazrafkan S, Corcoran P. Deep Learning for Consumer Devices and Services: Pushing the limits for machine learning, artificial intelligence, and computer vision. IEEE Consumer Electron. Mag 2017 Apr;6(2):48-56. [doi: 10.1109/MCE.2016.2640698]
143. Raza K. Artificial Intelligence Against COVID-19: A Meta-analysis of Current Research. In: Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach. Cham: Springer; Oct 13, 2020:165-176.

XSL•FO
RenderX

144. Ozsahin I, Sekeroglu B, Musa MS, Mustapha MT, Uzun Ozsahin D. Review on Diagnosis of COVID-19 from Chest CT Images Using Artificial Intelligence. Comput Math Methods Med 2020;2020:9756518 [FREE Full text] [doi: 10.1155/2020/9756518] [Medline: 33014121]
145. Zeeshan Hameed B, Patil V, Shetty D, Naik N, Nagaraj N, Sharma D. Use of Artificial Intelligence-based Computer Vision System to Practice Social Distancing in Hospitals to Prevent Transmission of COVID-19. Indian J Community Med 2020;45(3):379-380 [FREE Full text] [doi: 10.4103/ijcm.IJCM_366_20] [Medline: 33354024]
146. Loey M, Manogaran G, Taha M, Khalifa N. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement (Lond) 2021 Jan 01;167:108288 [FREE Full text] [doi: 10.1016/j.measurement.2020.108288] [Medline: 32834324]
147. Ud Din N, Javed K, Bae S, Yi J. A Novel GAN-Based Network for Unmasking of Masked Face. IEEE Access 2020;8:44276-44287. [doi: 10.1109/ACCESS.2020.2977386]

Abbreviations

AI: artificial intelligence

Edited by G Eysenbach, C Basch; submitted 27.11.20; peer-reviewed by I Ahmed, H Rai, L Wang, F Jijrees, LT Sen; comments to author 13.01.21; revised version received 21.01.21; accepted 07.04.21; published 25.05.21.

Please cite as:
Su Z, McDonnell D, Bentley BL, He J, Shi F, Cheshmehzangi A, Ahmad J, Jia P
Addressing Biodisaster X Threats With Artificial Intelligence and 6G Technologies: Literature Review and Critical Insights
J Med Internet Res 2021;23(5):e26109
URL: https://www.jmir.org/2021/5/e26109
doi:10.2196/26109
PMID:33961583

©Zhaohui Su, Dean McDonnell, Barry L Bentley, Jiguang He, Feng Shi, Ali Cheshmehzangi, Junaid Ahmad, Peng Jia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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

Personalization of Conversational Agent-Patient Interaction Styles for Chronic Disease Management: Two Consecutive Cross-sectional Questionnaire Studies

Christoph Gross[1,2*], BA, MA, PhD; Theresa Schachner[1,2*], BA, BSc, MSc, PhD; Andrea Hasl[3,4], BSc, MSc; Dario Kohlbrenner[5,6], MSc; Christian F Clarenbach[5], MD; Forian V Wangenheim[2], Prof Dr; Tobias Kowatsch[2,7,8], MSc, PhD, Dipl Inform (FH)

[1]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[3]Department of Educational Sciences, University of Potsdam, Potsdam, Germany

[4]International Max Planck Research School on the Life Course, Berlin, Germany

[5]Department of Pulmonology, University Hospital Zurich, Zurich, Switzerland

[6]Faculty of Medicine, University of Zurich, Zurich, Switzerland

[7]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[8]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Theresa Schachner, BA, BSc, MSc, PhD
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8006
Switzerland
Phone: 41 44 632 03 67
Email: tschachner@ethz.ch

Abstract

Background: Conversational agents (CAs) for chronic disease management are receiving increasing attention in academia and the industry. However, long-term adherence to CAs is still a challenge and needs to be explored. Personalization of CAs has the potential to improve long-term adherence and, with it, user satisfaction, task efficiency, perceived benefits, and intended behavior change. Research on personalized CAs has already addressed different aspects, such as personalized recommendations and anthropomorphic cues. However, detailed information on interaction styles between patients and CAs in the role of medical health care professionals is scant. Such interaction styles play essential roles for patient satisfaction, treatment adherence, and outcome, as has been shown for physician-patient interactions. Currently, it is not clear (1) whether chronically ill patients prefer a CA with a paternalistic, informative, interpretive, or deliberative interaction style, and (2) which factors influence these preferences.

Objective: We aimed to investigate the preferences of chronically ill patients for CA-delivered interaction styles.

Methods: We conducted two studies. The first study included a paper-based approach and explored the preferences of chronic obstructive pulmonary disease (COPD) patients for paternalistic, informative, interpretive, and deliberative CA-delivered interaction styles. Based on these results, a second study assessed the effects of the paternalistic and deliberative interaction styles on the relationship quality between the CA and patients via hierarchical multiple linear regression analyses in an online experiment with COPD patients. Patients' sociodemographic and disease-specific characteristics served as moderator variables.

Results: Study 1 with 117 COPD patients revealed a preference for the deliberative (50/117) and informative (34/117) interaction styles across demographic characteristics. All patients who preferred the paternalistic style over the other interaction styles had more severe COPD (three patients, Global Initiative for Chronic Obstructive Lung Disease class 3 or 4). In Study 2 with 123 newly recruited COPD patients, younger participants and participants with a less recent COPD diagnosis scored higher on interaction-related outcomes when interacting with a CA that delivered the deliberative interaction style (interaction between age and CA type: *relationship quality: b*=−0.77, 95% CI −1.37 to −0.18; *intention to continue interaction: b*=−0.49, 95% CI −0.97

XSL•FO
RenderX

to −0.01; *working alliance attachment bond: b*=−0.65, 95% CI −1.26 to −0.04; *working alliance goal agreement: b*=−0.59, 95% CI −1.18 to −0.01; interaction between recency of COPD diagnosis and CA type: *working alliance goal agreement: b*=0.57, 95% CI 0.01 to 1.13).

Conclusions: Our results indicate that age and a patient's personal disease experience inform which CA interaction style the patient should be paired with to achieve increased interaction-related outcomes with the CA. These results allow the design of personalized health care CAs with the goal to increase long-term adherence to health-promoting behavior.

(J Med Internet Res 2021;23(5):e26643) doi:10.2196/26643

KEYWORDS

conversational agents; chatbots; human-computer interaction; physician-patient interaction styles; deliberative interaction; paternalistic interaction; digital health; chronic conditions; disease management; COPD; chronic obstructive pulmonary disease

Introduction

The occurrence of chronic diseases is on the rise owing to greater longevity of the population, increasing exposure to environmental pollution, and unhealthy lifestyles [1]. As chronic diseases are not curable, related care is directed toward improving the functional status, reducing distressing symptoms, extending life duration through secondary prevention, and improving health-related quality of life [2,3]. This requires comprehensive and personalized disease management based on active long-term collaboration between health care practitioners and chronically ill patients [1].

However, disease management is time-consuming and staff-intensive and is thus often not sufficiently provided [1]. Conversational agents (CAs) (ie, computer programs that imitate interactions with humans) have the potential to improve the status quo as they allow for cheaper and scalable patient support outside the clinical setting [4,5]. When deployed on a smartphone, CAs remain easily accessible and can accompany patients in their daily lives [6,7]. However, long-term adherence to interventions delivered by health care CAs and the effectiveness of interventions with regard to health-related outcomes remain challenging [8,9].

To increase adherence and user value with respect to satisfaction, task efficiency, or the likelihood of sustained outcomes, personalization of CAs is viewed as promising [10]. Despite numerous design considerations [9] for health care CAs, such as personalized recommendations [10] and anthropomorphic cues [11], it is still unclear which CA-delivered interaction styles chronically ill patients prefer and whether the preference has an impact on CA-related perceptions (eg, working alliance) and health outcomes (eg, change in health-promoting behaviors). Research has singled out the importance of the interaction style for treatment satisfaction, adherence, and subsequent outcome [12-14] in face-to-face encounters between physicians and patients and in distance therapy via the phone, internet, or other means [15]. As people apply social behavior and expectations to computers or other media in the presence of anthropomorphic cues ("computers are social actors" paradigm) [16], CA-delivered interaction styles are expected to be of high relevance.

This paper applies and investigates the following four interaction styles of health care CAs [17]: (1) paternalistic (the physician, as a guardian [17], decides alone about the most appropriate treatment based on the assumption of shared values); (2) informative (the physician, as an expert [17], neutrally provides the patient with all treatment-related facts, so that the patient can choose); (3) interpretive (the physician, as a counsellor [17], helps the patient to elucidate the preferences and then leaves it to the patient to make a decision); and (4) deliberative (the physician, as a teacher or friend [17], conjointly discusses with the patient the best way forward).

Contemporary medical research advocates the deliberative style [18,19], which can also be referred to as shared decision making [20], as it is thought to consider patients' values and autonomy and the physician's caring role better than other interaction styles [17,20]. It was also the preferred interaction style by the majority of patients in preference studies [21,22]. However, there is evidence in the literature that sociodemographic and disease-related variables have an impact on the preferred interaction style. Older patients, for instance, tend to prefer a paternalistic interaction style [23,24], based on the assumption that they are accustomed to physicians being traditionally seen as an authority figure [25]. Among men, there is also a preference for the paternalistic interaction style [23,24]. Fatigue, lacking expertise or knowledge about the condition, and the fear of making a wrong decision are additional reasons mentioned in the literature that explain patients' preferences for a paternalistic interaction style in case of a severe condition, a newly diagnosed disease, or minor health literacy [26]. No influence of socioeconomic variables has been found [27] that could explain a preference for the informative style over the deliberative style. There seems to be no further evidence in the current literature base that talks about preferences for the interpretive interaction style.

To address these issues, we conducted two studies. The first study aimed to explore if there exist patient preferences for a paternalistic, informative, interpretive, or deliberative interaction style when a CA takes the role of a caregiver. The results of this study informed the second study that explored in more detail (1) which variables moderate preferences for the CA interaction style and (2) whether preferences have an impact on CA-related perceptions (eg, working alliance) and health outcomes (eg, change in health-promoting behaviors). Both studies involved patients diagnosed with chronic obstructive pulmonary disease (COPD), one of the global top four leading causes of premature death from chronic diseases [28].

XSL•FO
RenderX

Methods

Study Design

First, we conducted Study 1, a paper-pencil survey with COPD patients treated at a leading Swiss Hospital in the German-speaking part of Switzerland. Besides covering sociodemographic and health-related questions, the survey explored baseline differences in patient preferences for a deliberative, informative, interpretive, or paternalistic interaction style with a hypothetical health care CA.

The outcomes informed Study 2, an online experiment. For this study, we recruited COPD patients from four hospitals in the German-speaking part of Switzerland, from the Swiss Lung Association, and from an honorary led self-help association for COPD patients in the German-speaking part of Switzerland. We designed a between-subject online experiment where patients were randomly assigned to interact with a CA that followed either a deliberative or paternalistic interaction style. We chose these two styles since (1) we have already developed and experimentally tested the implementation of deliberative and paternalistic CA interactions in a recent study [29]; (2) there is the most information in the literature for these two styles regarding the moderating influences of sociodemographic and health-related variables; and (3) we expected to find significant effects when choosing the most and least preferred interaction styles as determined in Study 1. Both studies did not fall within the scope of Human Research Law, according to the local Swiss ethics authority, and thus did not require any formal authorization.

Sample Size Considerations

The primary objective of Study 1 was to explore whether general differences exist between interaction style preferences of COPD patients for their interaction with a CA. Thus, this part of the study was exploratory by nature and did not contain a detailed power analysis.

We conducted an a priori power analysis for Study 2 using R software (version 3.5.2) and the R package WebPower [30]. To identify a medium effect (f^2=0.15) [31] in a hierarchical multiple regression with an alpha level of .05, a statistical power of 0.80, a reduced model with one predictor, and a full model with 13 predictors, a total of 127 participants was required.

Inclusion Criteria

For Study 1, we defined the following inclusion criteria: (1) COPD diagnosis, (2) age of 18 years or older, and (3) ability to speak German.

We defined the same inclusion criteria for Study 2. Here, the first inclusion criterion was checked before distributing the link to the online experiment. The link was only sent out to patients who were registered as COPD patients at any of the participating hospitals, the lung association, or the self-help association. In addition, patients also had to confirm that they have COPD during the online experiment. The second and third inclusion criteria were checked at the beginning of the online experiment. If patients did not confirm either being of age or being able to speak German, the experiment was automatically completed. There were no exclusion criteria.

Procedure of Study 1 and Study 2

Study 1 was administered as a paper-and-pencil survey and divided into four parts. Before starting with the actual questions, we provided general information about the survey (ie, we clarified the objective and structure, and provided an illustrative explanation of a CA-based intervention). After querying sociodemographic and health-related questions, the survey explored patients' preferences regarding their interaction style with a hypothetical CA. Patients could choose from snippets portraying exemplary interactions with a CA, and each snippet depicted a different interaction style (see Figure 1 for an overview).

The procedure of Study 2 was as follows. Participants agreed to the study conditions and confirmed the study inclusion criteria (ie, being of age and being able to speak German). After querying standard demographic data (age, gender, mother tongue, and education), patients answered questions about their general health status and COPD. Patients were then randomly assigned to interact with a CA presenting a deliberative or paternalistic interaction style. The interactions were text-based and followed by a prescripted dialogue based on two predeveloped scripts. These scripts were developed and assessed in a recent study, where they were verified to be perceived as eliciting a deliberative or paternalistic interaction style between a CA and users [29]. During the conversation with the CA, patients could choose between one to three predefined answer options, which were identical in both conditions (ie, deliberative and paternalistic interaction styles). Deviations between the answer options only occurred when needed to keep the conversational flow realistic. Both interactions with the CA (ie, paternalistic and deliberative) were approximately of the same length and duration (38 conversational turns in the deliberative version and 32 in the paternalistic version with a reading duration of roughly over 12 minutes). After the interaction with the CA, participants were asked to evaluate the interaction with the CA on several dimensions. Details about the measures can be found in the measures section below.

To conclude the experiment, participants had to answer a short COPD health literacy quiz. The questions were based on the standardized Bristol COPD Knowledge Questionnaire (BCKQ), a multiple-choice questionnaire developed to measure the disease-specific knowledge of COPD patients [32]. Finally, patients reported on their perception of the length of the study and could leave some free-text feedback. All questions of Study 2 can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Illustrations outlining the four physician-patient interaction styles.



Technical Implementation of CAs for Study 2

We used Qualtrics, a software-based online survey and data collection platform, for the online experiment and for randomly assigning patients to one of the two experimental settings (ie, interaction with a CA using a deliberative interaction style and interaction with a CA using a paternalistic interaction style). We further used Collect.chat, a commercially available chatbot software, to develop the CA dialogues, and iframe to embed the CA into the Qualtrics HTML.

Recruitment and Management of Study Participants

We recruited the participants for Study 1 from the pulmonology department of a leading Swiss hospital. The patient database included around 1300 patients, and we contacted all by mail. Eligible patients received the printed survey and a letter containing information about the survey and participation conditions. We also provided a prefranked return envelope to reduce the necessary effort to reply to the survey and to minimize financial expenses for the participants. The postal send out took place on April 9 and 10, 2019, and we started to receive responses 1 week later. We received replies for 2 months in total, and the majority of replies reached us in the first 3 weeks.

The patients of Study 2 were recruited during a 3-month period from February to April 2020 at six study sites in Switzerland. The study sites were the pulmonology departments of four hospitals in the German-speaking part of Switzerland, the Swiss Lung Association, and an honorary-led self-help association for COPD patients. Participants were recruited via email or postal mail or in-person by participating health care professionals on-site. When they were recruited in-person, participating health care professionals handed out a flyer to potential participants. This flyer contained some information on the study and a link to the online experiment (see Multimedia Appendix 2 for the flyer). Participating health care professionals had previously received study instructions from the study authors. When patients were recruited via email, they received an email from the participating hospital, the Swiss Lung Association, or the self-help association. In this email, the participating contact person explained the study and asked for voluntary participation. The email further contained the same flyer as used for the on-site recruitment. When participants were recruited via postal mail, they received a letter from the participating hospital with study information and a link to the online experiment, which they had to type into a web browser. The postal letter is presented in Multimedia Appendix 3. We followed a multichannel and multisite recruiting process. More precisely, we contacted 903 patients at three study sites via email and postal mail. Additionally, 110 flyers were printed and displayed at two further study sites. Here, the flyers were displayed in the study sites' waiting rooms and were free to pick up for the patients. There were no additional prompts by medical personnel to participate in the study. The email and

XSL•FO
RenderX

postal mail had the same design and included exactly the same content as the flyer. At the sixth study site, 27 local COPD patient groups were contacted via email. Importantly, the resulting response rate has to be interpreted as a lower bound estimate since patients might have received an invitation more than once. For example, they could have been regular patients in hospital X and, at the same time, also a member of a COPD patient organization that distributed study invitations.

To start the online experiment, participants had to either click on the study link in the email or type the URL in their browser in case they received a postal letter. A simplified URL (ie, only portraying a shorter number of signs) was created with Bitly [33] to access the online experiment and to reduce barriers to participate. The experiment was available online from April to August 2020.

Measures of Study 1: Paper-and-Pencil Survey

For Study 1, we gathered basic sociodemographic (age, gender, and education) and health-related data (GOLD COPD level) [34]. GOLD stands for Global Initiative for Chronic Obstructive Lung Disease and is an internationally used scale for classifying the severity of COPD [34]. Patients reported sociodemographic and health-related data before being presented with the four main physician-patient interaction styles.

Sociodemographic Data

For the item *age*, patients reported their birth year. To derive patients' actual age, their birth year was subtracted from 2020 in the descriptive analyses. For the item *gender*, patients could choose their gender from a tick box with answer options "female," "male," and "no answer." For the item *education*, patients could choose one of the following four different options: *no formal education*, *secondary 1 education*, *secondary 2 education*, and *tertiary education*. Drawing on the 2011 International Standard Classification of Education (ISCED) Scheme [35], a comprehensive framework for comparing and organizing education programs and qualification across countries, these options translate to other global educational systems, such as the US system as follows. *Secondary 1 education* in Switzerland corresponds to level 2 of the ISCED scheme (ISCED 2, lower secondary education) and 9 years of education completed. ISCED 2 translates to completing junior high school in the United States. *Secondary 2 education* corresponds to level 3 of the ISCED scheme (ISCED 3, upper secondary education) and 13 years of education completed. ISCED 3 translates to completing a senior high school degree in the United States. *Tertiary education* in Switzerland (ISCED 6-8) encompasses bachelor's, master's, and PhD degrees, corresponding to a total of 16, 18, and 21 years needed to complete these educational levels (equivalent to the degrees in the United States) [36,37].

Health-Related Data

We queried the *GOLD COPD level* for each patient with the item "What is your GOLD rating?" (translation by the authors) on a 5-point scale ranging from 1 (GOLD 1) to 5 (I do not know). The highest and most severe GOLD level is GOLD 4.

Next, we introduced the four main physician-patient interaction styles as per a previous report [17], and patients were asked which of the four they would prefer for their interaction with a CA. We provided a definition of the respective role of the physician in the interaction style.

Measures of Study 2: Online Experiment

Study 2 was pretested with two pulmonologists of a leading Swiss hospital (both specialized in COPD), one advanced practitioner nurse in lung disease, and 10 PhD students of the Swiss Federal Institute of Technology in Zurich. We aimed to identify possible problems in terms of clarity, accuracy, and relevance for assessing health-related measures that are specific for COPD. Based on the feedback, a few changes were made to improve the wording of the questions and the order of the items.

For the main analyses and for assessing the various aspects of the two interaction styles, we gathered basic sociodemographic data (age, gender, and education), health-related data (GOLD COPD level, years since COPD diagnosis, and disease literacy), and interaction-related data (willingness to change, relationship quality, intention to continue interacting, and working alliance).

Sociodemographic and Health-Related Data

Sociodemographic data were gathered before the interaction with the CA. For the item *age*, patients could choose their birth year from a dropdown menu, ranging from 1900 to 2020. To derive patients' age, their birth year was subtracted from 2020 in the analyses. For the item *gender*, patients could choose their gender from a tick box with answer options "female," "male," and "no answer." For the item *education*, patients could choose from 12 different options from a dropdown menu (eg, "apprenticeship" [translated by the authors]; all original German measures can be found in Multimedia Appendix 4). For the analyses, educational attainment was recoded as years of education based on the 2011 ISCED scheme [35].

Health-related data were gathered before the interaction with the CA. Just as in Study 1, we first queried the *GOLD COPD level* for each patient with the item "What is your GOLD rating?" (translation by the authors) on a 5-point scale ranging from 1 (GOLD 1) to 5 (I do not know). The highest and most severe GOLD level is GOLD 4. Second, we measured *years since COPD diagnosis* with the item "Have you been diagnosed with COPD?" (translation by the authors), with answer options "yes" and "no" to ensure that patients confirm their COPD disease, and the item "If so, in which year?" (translation by the authors), with a dropdown menu to select a year between 1900 and 2020. To derive patients' years of experience with the disease, the diagnosis year was subtracted from 2020 in the analyses. Third, we measured *disease literacy*. Disease literacy was assessed using the BCKQ [32]. To keep the handling time for patients as short as possible, we selected 10 items of the BCKQ after consultation with two nurses who frequently use the questionnaire themselves with their patients. For example, patients had to mark the statement "COPD is commonly an inherited disease" (translation by the authors) as "true," "false," or "I do not know." If they evaluated the statement correctly, they were given 1 point, and in case of a false or "I do not know"

XSL•FO
RenderX

answer, they were given 0 points. We built a sum score over the 10 items, with higher values indicating higher disease literacy. The complete BCKQ (68 items) was reported to have reasonable internal consistency with a Cronbach alpha of .73 in previous studies [32]. The 10-item subset used in our study had a Cronbach alpha of .62, which, though not ideal, can still be considered acceptable [38], especially given the drastically reduced number of items.

Interaction-Related Data

Interaction-related data were gathered after the interaction with the CA Robo. We first measured *willingness to change* with the item "Was Robo able to motivate you for the proposed exercise?" (translation by the authors), with answer options "yes" and "no." This item was adapted from a previous report [39]. We further measured *relationship quality* with the German adapted item [40] "How would you characterize your relationship with Robo?" (translation by the authors), with answers on a scale ranging from 1 ("complete stranger") to 5 ("close friend"), and the German adapted item [41] "I think Robo liked me" (translation by the authors), with answers fixed on a 5-point Likert scale ranging from 1 ("not at all") to 5 ("very much"). Next, we measured *intention to continue interacting* with the German adapted item [42] "I would like to continue using Robo" (translation by the authors), with answers fixed on a 6-point Likert scale ranging from 1 ("strongly disagree") to 6 ("strongly agree"). Finally, we measured *working alliance* between the patients and the CA Robo with a German-adapted version of the Working Alliance Inventory-Short Revised (WAI-SR) [43]. Here, we measured the two subscales *attachment bond* and *goal agreement* (eg, "Robo and I respect each other," translation by the authors), with answers fixed on a 5-point Likert scale ranging from 1 ("rarely") to 5 ("always") (Multimedia Appendix 4). We decided to omit measuring the third subscale *task agreement* as it does not directly relate to our research questions. For secondary analyses (Multimedia Appendix 5), we measured additional interaction-related data (eg, satisfaction with the interaction and recommendation to a friend).

Results

Results of Study 1: Paper-and-Pencil Survey

Preferences for the paternalistic, informative, deliberative, or interpretive style are depicted by gender, age, educational levels, and disease severity. Out of 181 participants who started the survey, only 139 participants who completed the CA preference task were included in the final sample of Study 1. Moreover, 22 additional participants were excluded because they reported not having COPD. The final sample consisted of 117 participants with a mean age of 65.7 years and a mean GOLD classification value of 2.87. Of the 117 participants, 66 were male. Descriptive statistics can be found in Table 1. R scripts for all tables can be found in Multimedia Appendix 6. Missing data were dealt with by list-wise deletion because of the small number of participants having missing values for a variable of interest.

Table 1. Descriptive statistics of the participants in Study 1 (N=117).

Characteristic	Value
Gender, n (%)	
Male	66 (56%)
Female	51 (44%)
Age (years), mean (SD)	65.67 (10.92)
Education, n (%)	
None	7 (6%)
Secondary I	15 (13%)
Secondary II	56 (48%)
Tertiary	36 (31%)
COPD[a] severity value, mean (SD)	2.87 (0.98)
COPD severity, n (%)	
GOLD[b] 1	7 (6%)
GOLD 2	20 (17%)
GOLD 3	24 (21%)
GOLD 4	24 (21%)
Do not know	35 (30%)

[a]COPD: chronic obstructive pulmonary disease.

[b]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

XSL•FO
RenderX

Gender

Across CA categories, women most often chose the deliberative CA type, whereas men preferred the informative and deliberative CA types (26 vs 24; Table 2). Within each category of the deliberative, paternalistic, or interpretive CA type, men and women were fairly equally represented. Men constituted 77% (26/34 persons) of persons who preferred the informative CA type.

Table 2. Conversational agent preferences by gender.

Variable (gender)	Value, n[a] (%[b] by conversational agent category)				Sum (N=117), n
	Paternalistic (n=7)	Informative (n=34)	Interpretive (n=26)	Deliberative (n=50)	
Male	4 (57%)	26 (77%)	12 (46%)	24 (48%)	66
Female	3 (43%)	8 (24%)	14 (54%)	26 (52%)	51

[a]Numbers represent the absolute numbers of participants.

[b]Percentages of male/female participants present in each conversational agent category are given in parentheses.

Age

Younger participants (40-50 years old) preferred the deliberative CA type over the paternalistic CA type. Participants in the age group 51-60 years preferred the informative type. Participants in the age groups 61-70 and 71-80 years most often chose the deliberative type, and the oldest participants (81-90 years old) were fairly equally distributed across the informative, interpretive, and deliberative CA types. Within categories, 57% (4/7) of participants who chose the paternalistic CA type were in the youngest age category (40-50 years old). On the other hand, in the interpretive and deliberative groups, 78% (20/26) and 70% (35/50) of participants, respectively, were older than 60 years (Table 3).

Table 3. Conversational agent preferences by age.

Variable (age)	Value, n[a] (%[b] by conversational agent category)				Sum (N=117), n
	Paternalistic (n=7)	Informative (n=34)	Interpretive (n=26)	Deliberative (n=50)	
40-50 years	4 (57%)	1 (3%)	1 (4%)	6 (12%)	12
51-60 years	2 (29%)	12 (35%)	5 (19%)	9 (18%)	28
61-70 years	0 (0%)	9 (27%)	9 (35%)	14 (28%)	32
71-80 years	1 (14%)	9 (27%)	9 (35%)	19 (38%)	38
81-90 years	0 (0%)	3 (9%)	2 (8%)	2 (4%)	7

[a]Numbers represent the absolute numbers of participants.

[b]Percentages of age category participants present in each conversational agent category are given in parentheses.

Educational Levels

Participants without any formal education preferred the informative CA type. Participants who finished secondary I were fairly equally distributed across categories. Participants with higher educational levels (secondary II and tertiary) preferred the deliberative CA type (Table 4). This pattern was more distinct for secondary II than tertiary, where participants also often chose the informative CA type. Within categories, participants who had secondary II and tertiary education constituted 85% (40/47) of those who chose the deliberative type. Similarly, participants who had secondary II and tertiary education constituted 87% (23/26) of those who preferred the interpretive type. In the paternalistic condition, only the first three educational levels were represented (no formal education, secondary I, and secondary II).

Table 4. Conversational agent preferences by educational level.

Variable (education)	Value, n[a] (%[b] by conversational agent category)				Sum (N=114), n
	Paternalistic (n=7)	Informative (n=34)	Interpretive (n=26)	Deliberative (n=47)	
Secondary I	3 (43%)	4 (12%)	3 (12%)	5 (11%)	15
Secondary II	3 (43%)	14 (41%)	14 (54%)	25 (53%)	56
Tertiary	0 (0%)	12 (35%)	9 (35%)	15 (32%)	36
No formal education	1 (14%)	4 (12%)	0 (0%)	2 (4%)	7

[a]Numbers represent the absolute numbers of participants.

[b]Percentages of educational level participants present in each conversational agent category are given in parentheses.

XSL•FO
RenderX

Severity of COPD

Across CA categories, participants with less severe disease (GOLD 1) preferred the informative and interpretive CA types. Participants with mid-severe disease (GOLD 2 and 3) preferred the deliberative CA type. Participants with severe disease (GOLD 4) preferred the deliberative and informative CA types. Within categories, participants with higher disease levels (GOLD 3+4) constituted 100% (3/3) of those who preferred the paternalistic CA type (Table 5).

Table 5. Conversational agent preferences by Global Initiative for Chronic Obstructive Lung Disease classification.

Variable (GOLD[a] classification)	Value, n[b] (%[c] by conversational agent category)[d]				Sum (N=75), n
	Paternalistic (n=3)	Informative (n=27)	Interpretive (n=12)	Deliberative (n=33)	
GOLD 1	0 (0%)	3 (11%)	3 (25%)	1 (3%)	7
GOLD 2	0 (0%)	7 (26%)	4 (33%)	9 (27%)	20
GOLD 3	2 (67%)	6 (22%)	3 (25%)	13 (39%)	24
GOLD 4	1 (33%)	11 (41%)	2 (17%)	10 (30%)	24

[a]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[b]Numbers represent absolute numbers of participants.

[c]Percentages of GOLD classification category participants present in each conversational agent category are given in parentheses.

[d]The data of a maximum of seven out of 117 participants were deleted for Table 5 because they did not report any information on their GOLD classification (neither "GOLD 1-4" nor "I do not know").

Results of Study 2: Online Experiment

We conducted hierarchical multiple regression modeling to predict participants' self-reported interaction quality with the CA, that is, *willingness to change*, *relationship quality*, *intention to continue interaction*, and *working alliance*, based on (1) the type of CA (paternalistic/deliberative), (2) patients' demographics (age, gender, and education), and (3) COPD-related measures (GOLD, COPD disease literacy, and experience with COPD). Each outcome was predicted in a three-step procedure. The first block added to the model was the CA type (labelled "model 1"). The second block contained the CA type and participants' demographics (labelled "model 2"), and the third block consisted of the CA type, participants' demographics, and COPD-related measures (labelled "model 3"). *Relationship quality*, *intention to continue interaction*, and *working alliance* were measured on a metric scale. We calculated hierarchical linear regressions for those outcomes and logistic regression for the binary outcome *willingness to change*. As in Study 1, missing data were dealt with by list-wise deletion.

Analysis of Sociodemographic and Health-Related Data

The descriptive statistics of the experiment are shown in Table 6. Out of 168 participants who started the survey, 124 completed the survey. One additional participant was excluded because of age (<18 years old), leading to a final sample of 123 participants. Of those 123 participants, 76 were male, with a mean age of 67.8 years and a mean duration of 8.4 years since their COPD diagnosis. The mean GOLD classification value was 2.70.

Table 6. Descriptive statistics of the participants in Study 2 (N=123).

Characteristic	Value
Gender, n (%)	
Male	76 (62%)
Female	47 (38%)
Age (years), mean (SD)	67.82 (9.37)
Education (years of formal education), mean (SD)	14.28 (2.37)
COPD[a] severity value, mean (SD)	2.70 (0.88)
COPD severity, n (%)	
GOLD[b] 1	6 (5%)
GOLD 2	24 (20%)
GOLD 3	29 (24%)
GOLD 4	14 (11%)
Do not know	35 (28%)

[a]COPD: chronic obstructive pulmonary disease.

[b]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

XSL•FO
RenderX

Analyses of Interaction-Related Data

For our analysis, we introduce interaction-related outcomes, defined in terms of the outcome variables *willingness to change*, *relationship quality*, *intention to continue interaction*, and two dimensions of *working alliance* (*attachment bond* and *goal agreement*). Better interaction-related outcomes indicate a higher willingness to change one's behavior after interaction with the CA, a higher perceived relationship quality, a higher motivation to continue interacting with the CA, and a higher-rated reported working alliance with the CA in terms of perceived close attachment bond and common goal agreement. Multimedia Appendix 5 presents the results of additional interaction-related data (eg, satisfaction with the interaction and recommendation to a friend). The R scripts for all analyses are presented in Multimedia Appendix 7.

Willingness to Change

Overall, participants who interacted with a paternalistic CA reported being more willing to change their behavior based on the CA intervention than those who worked with a deliberative CA (Table 7). There were no substantial interaction effects between CA type and participants' demographics or CA type and patients' COPD-related characteristics.

Table 7. Regression of conversational agent type, participants' demographics, and chronic obstructive pulmonary disease–related characteristics in terms of participants' willingness to change their behavior after conversational agent interaction.

Variable	Model 1[a], regression coefficient (95% CI)	Model 2[b], regression coefficient (95% CI)	Model 3[c], regression coefficient (95% CI)
Intercept	0.833 (0.724 to 0.943)	0.840 (0.699 to 0.980)	0.840 (0.640 to 1.039)
CA[d] type[e]	−0.183 (−0.338 to −0.029)[f]	−0.179 (−0.388 to 0.030)	−0.074 (−0.350 to 0.203)
Gender[g]	N/A[h]	−0.024 (−0.283 to 0.235)	0.041 (−0.325 to 0.406)
Age	N/A	−0.037 (−0.161 to 0.088)	−0.010 (−0.207 to 0.188)
Education[i]	N/A	−0.108 (−0.229 to 0.012)	−0.064 (−0.231 to 0.103)
Gender*CA type	N/A	−0.035 (−0.387 to 0.317)	−0.163 (−0.627 to 0.301)
Age*CA type	N/A	0.003 (−0.162 to 0.169)	−0.145 (−0.386 to 0.095)
Education*CA type	N/A	0.091 (−0.078 to 0.259)	0.055 (−0.157 to 0.268)
GOLD[j]	N/A	N/A	0.053 (−0.105 to 0.210)
COPD[k] literacy	N/A	N/A	0.165 (−0.047 to 0.377)
Experience[l]	N/A	N/A	−0.127 (−0.294 to 0.040)
GOLD*CA type	N/A	N/A	−0.040 (−0.261 to 0.181)
COPD literacy*CA type	N/A	N/A	−0.058 (−0.311 to 0.194)
Experience*CA type	N/A	N/A	0.152 (−0.078 to 0.381)

[a]Model 1 includes the CA type. It has 120 observations and an Aikaike Information Criterion value of 142.880 (a smaller value is associated with better model fit).

[b]Model 2 includes the CA type and participants' demographics. It has 113 observations and an Aikaike Information Criterion value of 146.953 (a smaller value is associated with better model fit).

[c]Model 3 includes the CA type, participants' demographics, and COPD-related measures. It has 67 observations and an Aikaike Information Criterion value of 83.496 (a smaller value is associated with better model fit).

[d]CA: conversational agent.

[e]CA type is coded as follows: 0=paternalistic, 1=deliberative.

[f]Significant (P<.05).

[g]Gender is coded as follows: 0=male, 1=female.

[h]N/A: not applicable.

[i]Education is measured in years of formal education.

[j]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[k]COPD: chronic obstructive pulmonary disease.

[l]Experience with COPD in years since COPD diagnosis.

Relationship Quality

On average, older participants reported better relationship quality with the CA than younger participants, irrespective of the CA type. Participants with more severe COPD reported better relationship quality with the CA than participants with less severe COPD, irrespective of the CA type. There was a negative interaction effect between CA type and age, implying that older participants preferred a paternalistic CA and younger participants preferred a deliberative CA with respect to relationship quality (Table 8).

XSL•FO
RenderX

Table 8. Regression of conversational agent type, participants' demographics, and chronic obstructive pulmonary disease–related characteristics in terms of participants' relationship quality with the conversational agent.

Variable	Model 1[a], regression coefficient (95% CI)	Model 2[b], regression coefficient (95% CI)	Model 3[c], regression coefficient (95% CI)
Intercept	−0.073 (−0.326 to 0.181)	−0.030 (−0.350 to 0.290)	−0.284 (−0.774 to 0.206)
CA[d] type[e]	0.146 (−0.213 to 0.504)	−0.007 (−0.482 to 0.469)	0.176 (−0.504 to 0.857)
Gender[f]	N/A[g]	−0.399 (−0.988 to 0.191)	0.262 (−0.637 to 1.161)
Age	N/A	0.155 (−0.129 to 0.438)	0.615 (0.129 to 1.100)[h]
Education[i]	N/A	−0.031 (−0.306 to 0.243)	0.240 (−0.171 to 0.652)
Gender*CA type	N/A	0.546 (−0.255 to 1.347)	−0.377 (−1.519 to 0.764)
Age*CA type	N/A	−0.215 (−0.592 to 0.162)	−0.774 (−1.366 to −0.181)[h]
Education*CA type	N/A	0.057 (−0.326 to 0.440)	−0.018 (−0.541 to 0.504)
GOLD[j]	N/A	N/A	0.398 (0.011 to 0.786)[h]
COPD[k] literacy	N/A	N/A	0.324 (−0.198 to 0.846)
Experience[l]	N/A	N/A	−0.220 (−0.631 to 0.191)
GOLD*CA type	N/A	N/A	−0.132 (−0.675 to 0.411)
COPD literacy*CA Type	N/A	N/A	−0.031 (−0.652 to 0.590)
Experience*CA type	N/A	N/A	0.194 (−0.370 to 0.759)

[a]Model 1 includes the CA type. It has 120 observations and an R^2 value of 0.005.

[b]Model 2 includes the CA type and participants' demographics. It has 113 observations and an R^2 value of 0.039.

[c]Model 3 includes the CA type, participants' demographics, and COPD-related measures. It has 67 observations and an R^2 value of 0.283.

[d]CA: conversational agent.

[e]CA type is coded as follows: 0=paternalistic, 1=deliberative.

[f]Gender is coded as follows: 0=male, 1=female.

[g]N/A: not applicable.

[h]Significant (P<.05).

[i]Education is measured in years of formal education.

[j]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[k]COPD: chronic obstructive pulmonary disease.

[l]Experience with COPD in years since COPD diagnosis.

Intention to Continue Interaction

Disease severity positively predicted participants' intention to continue interacting with the CA after the interaction ended. The higher a participant's GOLD classification, the higher was his or her intention to continue (Table 9). Participants with fewer years of experience with COPD reported a higher intention to continue the interaction, irrespective of the assigned CA type. Older participants reported being more likely to continue the CA interaction when working with a paternalistic CA, and younger participants reported that when working with a deliberative CA.

XSL•FO
RenderX

Table 9. Regression of conversational agent type, participants' demographics, and chronic obstructive pulmonary disease–related characteristics in terms of participants' intention to continue interacting with the conversational agent.

Variable	Model 1[a], regression coefficient (95% CI)	Model 2[b], regression coefficient (95% CI)	Model 3[c], regression coefficient (95% CI)
Intercept	0.112 (−0.140 to 0.365)	0.046 (−0.277 to 0.369)	0.134 (−0.263 to 0.531)
CA[d] type[e]	−0.224 (−0.581 to 0.133)	−0.169 (−0.649 to 0.311)	−0.279 (−0.831 to 0.272)
Gender[f]	N/A[g]	−0.037 (−0.632 to 0.558)	0.404 (−0.325 to 1.133)
Age	N/A	−0.075 (−0.361 to 0.211)	0.178 (−0.216 to 0.572)
Education[h]	N/A	0.053 (−0.224 to 0.331)	0.019 (−0.314 to 0.353)
Gender*CA type	N/A	−0.011 (−0.819 to 0.797)	−0.637 (−1.563 to 0.288)
Age*CA type	N/A	−0.063 (−0.444 to 0.317)	−0.485 (−0.965 to −0.005)[i]
Education*CA type	N/A	−0.162 (−0.549 to 0.224)	−0.084 (−0.507 to 0.340)
GOLD[j]	N/A	N/A	0.420 (0.106 to 0.734)[i]
COPD[k] literacy	N/A	N/A	0.199 (−0.224 to 0.622)
Experience[l]	N/A	N/A	−0.391 (−0.724 to −0.058)[i]
GOLD*CA type	N/A	N/A	−0.153 (−0.593 to 0.287)
COPD literacy*CA type	N/A	N/A	0.277 (−0.226 to 0.781)
Experience*CA type	N/A	N/A	0.277 (−0.181 to 0.735)

[a]Model 1 includes the CA type. It has 120 observations and an R^2 value of 0.013.

[b]Model 2 includes the CA type and participants' demographics. It has 113 observations and an R^2 value of 0.026.

[c]Model 3 includes the CA type, participants' demographics, and COPD-related measures. It has 67 observations and an R^2 value of 0.411.

[d]CA: conversational agent.

[e]CA type is coded as follows: 0=paternalistic, 1=deliberative.

[f]Gender is coded as follows: 0=male, 1=female.

[g]N/A: not applicable.

[h]Education is measured in years of formal education.

[i]Significant (P<.05).

[j]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[k]COPD: chronic obstructive pulmonary disease.

[l]Experience with COPD in years since COPD diagnosis.

Working Alliance (Attachment Bond)

We found a substantial negative interaction effect among age, CA type, and reported attachment bond with the CA. This indicates that older participants had a higher attachment bond when working with the paternalistic CA type and younger participants had that when working with the deliberative CA type. Overall, participants who had a higher disease literacy of COPD also reported better attachment bond (Table 10), irrespective of the assigned CA type.

XSL•FO
RenderX

Table 10. Regression of conversational agent type, participants' demographics, and chronic obstructive pulmonary disease–related characteristics in terms of participants' working alliance (attachment bond) with the conversational agent.

Variable	Model 1[a], regression coefficient (95% CI)	Model 2[b], regression coefficient (95% CI)	Model 3[c], regression coefficient (95% CI)
Intercept	−0.123 (−0.375 to 0.129)	−0.161 (−0.482 to 0.161)	−0.424 (−0.932 to 0.083)
CA[d] type[e]	0.245 (−0.111 to 0.602)	0.339 (−0.138 to 0.817)	0.526 (−0.178 to 1.231)
Gender[f]	N/A[g]	0.001 (−0.591 to 0.593)	0.721 (−0.210 to 1.651)
Age	N/A	0.141 (−0.144 to 0.425)	0.302 (−0.200 to 0.805)
Education[h]	N/A	−0.094 (−0.370 to 0.181)	0.061 (−0.365 to 0.487)
Gender*CA type	N/A	−0.186 (−0.990 to 0.619)	−0.939 (−2.121 to 0.242)
Age*CA type	N/A	−0.284 (−0.662 to 0.095)	−0.650 (−1.263 to −0.037)[i]
Education*CA type	N/A	0.047 (−0.338 to 0.432)	−0.022 (−0.563 to 0.518)
GOLD[j]	N/A	N/A	0.278 (−0.123 to 0.679)
COPD[k] literacy	N/A	N/A	0.595 (0.055 to 1.135)[i]
Experience[l]	N/A	N/A	−0.307 (−0.732 to 0.118)
GOLD*CA type	N/A	N/A	−0.420 (−0.982 to 0.142)
COPD literacy*CA type	N/A	N/A	−0.405 (−1.048 to 0.237)
Experience*CA type	N/A	N/A	0.537 (−0.047 to 1.122)

[a]Model 1 includes the CA type. It has 120 observations and an R^2 value of 0.015.

[b]Model 2 includes the CA type and participants' demographics. It has 113 observations and an R^2 value of 0.044.

[c]Model 3 includes the CA type, participants' demographics, and COPD-related measures. It has 67 observations and an R^2 value of 0.221.

[d]CA: conversational agent.

[e]CA type is coded as follows: 0=paternalistic, 1=deliberative.

[f]Gender is coded as follows: 0=male, 1=female.

[g]N/A: not applicable.

[h]Education is measured in years of formal education.

[i]Significant (P<.05).

[j]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[k]COPD: chronic obstructive pulmonary disease.

[l]Experience with COPD in years since COPD diagnosis.

Working Alliance (Goal Agreement)

Irrespective of the CA type participants were working with, those with a higher disease literacy reported higher perceived support to achieve their goals by the CA. Participants with fewer years of experience with COPD reported higher perceived support to achieve their goals by the CA. Older participants reported higher support by the CA when in the paternalistic condition, and younger participants reported that when in the deliberative condition. A positive interaction effect between CA type and experience with COPD implied that participants who were more experienced with COPD reported better perceived support in achieving their goals when interacting with a deliberative CA and participants who were less experienced with COPD reported that when interacting with a paternalistic CA (Table 11).

XSL•FO
RenderX

Table 11. Regression of conversational agent type, participants' demographics, and chronic obstructive pulmonary disease–related characteristics in terms of participants' working alliance (goal agreement) with the conversational agent.

Variable	Model 1[a], regression coefficient (95% CI)	Model 2[b], regression coefficient (95% CI)	Model 3[c], regression coefficient (95% CI)
Intercept	0.017 (−0.237 to 0.271)	0.001 (−0.322 to 0.323)	−0.101 (−0.585 to 0.383)
CA[d] type[e]	−0.034 (−0.394 to 0.325)	0.120 (−0.358 to 0.599)	0.219 (−0.453 to 0.891)
Gender[f]	N/A[g]	−0.066 (−0.660 to 0.528)	0.515 (−0.373 to 1.402)
Age	N/A	−0.000 (−0.286 to 0.285)	0.247 (−0.233 to 0.726)
Education[h]	N/A	−0.126 (−0.402 to 0.151)	−0.201 (−0.608 to 0.206)
Gender*CA type	N/A	−0.256 (−1.063 to 0.551)	−0.840 (−1.967 to 0.287)
Age*CA type	N/A	−0.134 (−0.514 to 0.245)	−0.592 (−1.176 to −0.007)[i]
Education*CA type	N/A	0.158 (−0.228 to 0.544)	0.233 (−0.283 to 0.748)
GOLD[j]	N/A	N/A	0.163 (−0.219 to 0.546)
COPD[k] literacy	N/A	N/A	0.690 (0.175 to 1.205)[i]
Experience[l]	N/A	N/A	−0.435 (−0.841 to −0.029)[i]
GOLD*CA type	N/A	N/A	−0.324 (−0.860 to 0.212)
COPD literacy*CA type	N/A	N/A	−0.384 (−0.997 to 0.229)
Experience*CA type	N/A	N/A	0.570 (0.012 to 1.128)[i]

[a]Model 1 includes the CA type. It has 120 observations and an R^2 value of 0.0003.

[b]Model 2 includes the CA type and participants' demographics. It has 113 observations and an R^2 value of 0.030.

[c]Model 3 includes the CA type, participants' demographics, and COPD-related measures. It has 67 observations and an R^2 value of 0.273.

[d]CA: conversational agent.

[e]CA type is coded as follows: 0=paternalistic, 1=deliberative.

[f]Gender is coded as follows: 0=male, 1=female.

[g]N/A: not applicable.

[h]Education is measured in years of formal education.

[i]Significant (P<.05).

[j]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[k]COPD: chronic obstructive pulmonary disease.

[l]Experience with COPD in years since COPD diagnosis.

Summary

In summary, we found evidence that age and experience with COPD inform participants' preferences for a deliberative or paternalistic interaction style of the CA. Older participants reported better interaction-related outcomes when interacting with a paternalistic CA, whereas younger participants reported that when interacting with a deliberative CA. Participants with fewer years of personal experience with COPD reported better interaction-related outcomes when interacting with a paternalistic CA, whereas those with more years of personal experience reported that when interacting with a deliberative CA. We did not find evidence for gender, disease level, and disease literacy. Irrespective of the CA type, disease literacy positively predicted both dimensions of working alliance, and participants with fewer years of experience with COPD reported higher perceived support in goal agreement by the CA and were more motivated to continue the interaction with the CA. A more severe disease level was associated with higher motivation to continue the interaction with the CA. Participants who worked with a paternalistic CA were more likely to change their behavior based on the intervention. Thus, our results indicate that knowing the age and years of experience of a patient with COPD can help to decide which interaction style to choose for the patient in order to increase interaction-related outcomes for the patient.

Discussion

Principal Findings

In this work, we investigated the preferences of patients with COPD for specific interaction styles of health care CAs. The interaction style between health care professionals and patients has long been recognized as a key success factor for chronic disease management and final treatment success [44,45]. Given the rising number of chronically diseased patients and the associated financial and personal burdens, CAs represent scalable and ubiquitous digital tools to support chronic patients and relieve human health care professionals. A systematic approach for inducing two specific interaction styles into CAs

XSL•FO
RenderX

in a health care setting has previously been developed and validated [29].

In our first study, we determined baseline differences for preferred interaction styles between 117 COPD patients and CAs. We showed that differences in preferences for specific interaction styles for the interaction between chronically diseased patients and CAs exist. In our second study, we explored the patterns of preferences for two specific interaction styles in 123 COPD patients. We found evidence that younger patients reported better interaction-related outcomes when interacting with a deliberative CA, while older COPD patients reported better interaction-related outcomes when interacting with a paternalistic CA. Additionally, COPD patients with longer personal experience with the disease reported better interaction-related outcomes when interacting with a deliberative CA. Moreover, COPD patients with lower COPD disease literacy reported better interaction-related outcomes when interacting with a paternalistic CA. Gender, disease severity, or disease literacy did not affect any preferences for specific interaction styles. Nevertheless, we found evidence that disease literacy, in general, positively predicted both dimensions of working alliance independent of the interaction style.

This paper is especially important for the development of personalized CAs in the context of digital health care, with a focus on chronic diseases. To our knowledge, this is the first investigation that systematically evaluated the preferences of chronic patients for their interaction style with CAs. While CAs have primarily been developed portraying a single interaction style for every human counterpart interacting with them, medical research has long stated the crucial importance of deploying personal interaction styles in order to improve patient satisfaction [40], treatment adherence, and final treatment outcome [46,47]. Addressing the gap in the literature regarding differentiated and personalized interaction styles for patient-CA interactions and adding to the growing body of literature on CA personalization [10,11], this paper now provides the first evidence that chronic patients report better interaction-related outcomes when interacting with CAs that display personalized interaction styles.

The findings of these two studies further inform the pairing of chronic patients to CAs that are personalized at the level of their interaction style. While medical research postulates the relevance of five factors (gender [23,24], age [23,24], disease level [26,48], personal experience with a disease [49], and disease literacy [50,51]) that influence the patient-physician interaction, we showed that not all of these aspects are similarly important when it comes to coupling chronic patients with CAs. Our first results indicate that knowing the age and personal disease experience of the patient is sufficient to decide which interaction style results in increased interaction-related outcomes for the patient at hand. While these are the first results from a restricted sample in an experimental setting, the implications for CA deployment could be significant. Especially from a privacy perspective, the findings would reduce the amount of personal patient data needed to achieve an advantageous CA-patient allocation, as only these two data points can be gathered instead of obtaining a whole plethora of personal data. In addition, these two data points can be easily collected at the start of the patient-CA interaction, without any specific (medical) knowledge needed for assessing them. This could reduce the work of health care professionals, whose time is limited and costly, as the best-fitting CA could be allocated based on responses to a simple digital questionnaire at the beginning of the patient-CA interaction. Notwithstanding these potential possibilities, more research needs to be done to be able to robustly understand patients' preferences in detail so that industry applications can be developed and reliably used.

Strengths and Limitations

This paper has several strengths. First, we followed a two-step approach by determining baseline differences of COPD patients' preferences for their interaction with a CA in Study 1 and subsequently expanding the findings to Study 2, a between-subject online experiment. Second, we deployed a systematic and validated approach for inducing two specific interaction styles into the patient-CA interaction [29]. Third, we continuously ensured an objective approach by integrating both theoretical knowledge and applied medical expertise into the development of the experiment. We did this by closely collaborating with medical professionals. Here, we worked together with not only medical experts on COPD (the chronic disease subject of this paper), but also health care professionals from other fields to reduce the risk of bias. In addition, we integrated the views of both senior and novice health care professionals to reflect traditional paternalistic-based training and current shared decision making–based training. Fourth, we focused on investigating the preferences of a specific target population, that is, patients with COPD. This focus on a relatively homogeneous patient group allowed us to delve into depth and gain a profound understanding of their preferences.

This work also has limitations. First, we tested baseline differences between all four major interaction styles in Study 1 (a paper-pencil study), but we only tested personal preferences for the deliberative and paternalistic interaction styles in Study 2 (online experiment). These are the two interaction styles where a systematic and validated approach for inducing the specific interaction style into the CA-patient interaction for a digital online setting exists. They further represent the start point and endpoint of a hypothetical ethical development process of a model patient-physician interaction [52]. Nevertheless, the results from Study 1 indicated that some patients might have personal preferences for other interaction styles than these two. The preferences for these interaction styles need to be investigated by future research. In addition, we could see differences in the preference allocation between the studies. While older participants in Study 2 preferred a paternalistic CA when evaluating relationship quality, 57% (4/7) of participants who chose the paternalistic style in Study 1 were in the youngest of the tested age ranges. This difference could be assumed to have originated in the respective study setup. Participants could choose from one single interaction snippet of the four main different interaction styles, and participants in Study 2 had a real interaction with a CA, but only the deliberative or paternalistic CA. As such, further research is needed to robustly understand patients' preferences, especially when it comes to a real interaction with a CA. Second, the study population only included German-speaking patients based in one country. It

XSL•FO
RenderX

could be that other languages or regions influence different interaction style preferences and personal requirements. Third, disease-related patient inputs (eg, GOLD status and years of diagnosis) were self-reported and hence not verifiable. Additionally, not all patients reported their GOLD status. Fourth, we only modeled the first part of an initial interaction between a patient and a CA. In reality, patients would need to interact over a prolonged period of time when a CA supports their chronic disease management. Fifth, we used a paper-based snippet of a hypothetical patient-CA interaction in Study 1 and a prescripted and rule-based CA in Study 2. While these two approaches were necessary because of the study condition in Study 1 (we sent a physical letter to the patients) and to control the experimental condition in Study 2, both approaches have their limitations when it comes to emulating a naturalistic patient-physician interaction. We are aware of the increasing number of artificial intelligence (AI)–based CAs [4] as well as voice-based CAs for health care purposes [53]. We believe this could be an interesting path for future research in this context of personalized patient-CA interaction styles. AI-based CAs could not only interact in a more naturalistic way by utilizing unconstrained written, spoken, or visual input [4], but also further adapt dynamically to personal developments, for example, the level of disease literacy of their human users.

Suggestions for Future Research

In general, we advise future research to put a stronger focus on the investigation of patients' personal preferences for specific interaction styles when interacting with CAs based on the long-known importance of this factor in the human patient-human physician context. In detail, we see specific possibilities for future research motivated by the limitations of this study and as an extension of it.

First, we advise future research to expand and test the used systematic approach for inducing two specific interaction styles to more interaction styles. As discussed above in the Strengths and Limitations subsection, we could see variations in preferences when patients could choose from four interaction styles (but only having one written interaction snippet to choose from) and a setting where they could interact for longer with either a deliberative or paternalistic CA. This could provide valuable insights into how many different interaction styles for patient-CA interactions are needed. In addition, we recommend future research to study the development of patients' preferences over time. It would be highly relevant to determine whether such preferences stay stable or dynamically evolve over time. Second, we suggest the development and evaluation of CAs in other languages besides German and in more diverse geographical settings to investigate the effects of language and regional specificities on patient-CA interaction styles. Third, we recommend focusing on the preferences of patients suffering from different diseases (both acute and chronic). We suggest focusing on differences within as well as between diseases to understand any influencing factors of the medical condition at hand in detail. Fourth, future research could expand our experiment and develop a more extended interaction between patients and interaction style–personalized CAs. This could bear interesting findings to further understand dynamic developments of personal preferences for interaction styles between patients and CAs. Fifth, we believe the development and implementation of AI-based CAs that are able to interact more naturally and adapt dynamically to the patient at hand could yield interesting results in the field of patients' personal preferences for their interaction with CAs.

Conclusions

The interaction style between patients and physicians is recognized as a critical parameter for patient satisfaction, treatment adherence, and subsequent treatment outcome and, as such, also plays a paramount role for chronic disease management. So far, CAs as ubiquitous and scalable digital tools have mainly utilized a single interaction style for every patient, thus ignoring the relevance of personalized interaction styles. In this paper, we showed that chronically diseased patients exhibit preferences for different interaction styles when conversing with a digital health CA. Our results provide evidence that patients' age and personal experiences with the disease inform their preferences for a specific interaction style. Hereby, this work provides insights into the rising trend of personalized CAs in health care. We envisage a future where every chronic patient gets paired with a CA exhibiting the right interaction style at the right moment and dynamically adapting to the needs of the patient, thereby allowing for a satisfying and fulfilling patient-CA interaction that supports the best possible treatment outcomes and disease management.

Acknowledgments

We want to express our deepest gratitude to the hospitals across Switzerland, the Swiss Lung Association, and the self-help association for COPD patients (www.copd-sauerstoffpatienten.ch) that allowed us to recruit study participants from their patient population member base. Without them, a study of this scale would not have been possible. We would also like to thank our fellow researchers at ETH Zurich, who helped us pretest the online experiment to guarantee a seamless experience for the study participants. We would also like to express our gratitude to Marcia Nießen of ETH Zurich for her considerations to the questionnaire of Study 1.

Conflicts of Interest

FW co-chairs the Center for Digital Health Interventions (CDHI), a joint initiative between the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen. The CDHI is funded in part by the Swiss health insurer CSS. TK is a scientific director at the CDHI and a co-founder of Pathmate Technologies, a university spin-off company that delivers digital clinical pathways with the help of conversational agents. Neither

XSL•FO
RenderX

CSS nor Pathmate Technologies was involved in the study described in this paper in any way. None of the health care professionals mentioned in this study were involved in the data analysis in any way. CG, TS, AH, and DK report no conflicts of interest.

Multimedia Appendix 1
Online experiment in Study 2 (translated by the authors into English for this paper).
[PDF File (Adobe PDF File), 314 KB - jmir_v23i5e26643_app1.pdf]

Multimedia Appendix 2
Chronic obstructive pulmonary disease chatbot study flyer (translated by the authors into English for this paper).
[PDF File (Adobe PDF File), 361 KB - jmir_v23i5e26643_app2.pdf]

Multimedia Appendix 3
Chronic obstructive pulmonary disease chatbot study letter (translated by the authors into English for this paper).
[PDF File (Adobe PDF File), 214 KB - jmir_v23i5e26643_app3.pdf]

Multimedia Appendix 4
Online experiment in Study 2: Original German measures.
[PDF File (Adobe PDF File), 793 KB - jmir_v23i5e26643_app4.pdf]

Multimedia Appendix 5
Secondary analyses.
[DOCX File , 31 KB - jmir_v23i5e26643_app5.docx]

Multimedia Appendix 6
R scripts for Study 1.
[ZIP File (Zip Archive), 1 KB - jmir_v23i5e26643_app6.zip]

Multimedia Appendix 7
R scripts for Study 2.
[ZIP File (Zip Archive), 4 KB - jmir_v23i5e26643_app7.zip]

References

1. Milani RV, Lavie CJ. Health care 2020: reengineering health care delivery to combat chronic disease. Am J Med 2015 Apr;128(4):337-343. [doi: 10.1016/j.amjmed.2014.10.047] [Medline: 25460529]
2. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, Kasteren YV. Health Care Customer Value Cocreation Practice Styles. Journal of Service Research 2012 May 25;15(4):370-389. [doi: 10.1177/1094670512442806]
3. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar 10;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
4. Schachner T, Keller R, V Wangenheim F. Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
5. Herring SC. Slouching Toward the Ordinary: Current Trends in Computer-Mediated Communication. New Media & Society 2016 Jun 30;6(1):26-36. [doi: 10.1177/1461444804039906]
6. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun 10;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
7. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
8. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
9. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design Features of Embodied Conversational Agents in eHealth: a Literature Review. International Journal of Human-Computer Studies 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
10. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]

XSL•FO
RenderX

11. Kang S, Feng A, Leuski A, Casas D, Shapiro A. The effect of an animated virtual character on mobile chat interactions. In: HAI '15: Proceedings of the 3rd International Conference on Human-Agent Interaction. 2015 Oct Presented at: 3rd International Conference on Human-Agent Interaction; October 2015; Daegu, Korea p. 105-112. [doi: 10.1145/2814940.2814957]
12. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [doi: 10.1016/0277-9536(94)00155-m] [Medline: 7792630]
13. Kaplan SH, Greenfield S, Ware JE. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. Med Care 1989 Mar;27(3 Suppl):S110-S127. [doi: 10.1097/00005650-198903001-00010] [Medline: 2646486]
14. Okunrintemi V, Spatz ES, Di Capua P, Salami JA, Valero-Elizondo J, Warraich H, et al. Patient–Provider Communication and Health Outcomes Among Individuals With Atherosclerotic Cardiovascular Disease in the United States. Circ Cardiovasc Qual Outcomes 2017 Apr;10(4):1-8. [doi: 10.1161/circoutcomes.117.003635]
15. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
16. Nass C, Steuer J, Tauber E. Computers are social actors. In: CHI '94: Conference Companion on Human Factors in Computing Systems. 1994 Apr Presented at: 1994 Conference Companion on Human Factors in Computing Systems; April 24, 1994; Boston, MA p. 72-78. [doi: 10.1145/259963.260288]
17. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. JAMA 1992;267(16):2221-2226. [Medline: 1556799]
18. Heritage J, Maynard DW. Problems and Prospects in the Study of Physician-Patient Interaction: 30 Years of Research. Annu. Rev. Sociol 2006 Aug;32(1):351-374. [doi: 10.1146/annurev.soc.32.082905.093959]
19. Borza LR, Gavrilovici C, Stockman R. Ethical Models Of Physician: Patient Relationship Revisited With Regard To Patient Autonomy, Values And Patient Education. Rev Med Chir Soc Med Nat Iasi 2015;119(2):496-501. [Medline: 26204658]
20. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
21. Bieber C, Müller KG, Blumenstiel K, Schneider A, Richter A, Wilke S, et al. Long-term effects of a shared decision-making intervention on physician-patient interaction and outcome in fibromyalgia. A qualitative and quantitative 1 year follow-up of a randomized controlled trial. Patient Educ Couns 2006 Nov;63(3):357-366. [doi: 10.1016/j.pec.2006.05.003] [Medline: 16872795]
22. Deber RB, Kraetschmer N, Urowitz S, Sharpe N. Do people want to be autonomous patients? Preferred roles in treatment decision-making in several patient populations. Health Expect 2007 Sep;10(3):248-258 [FREE Full text] [doi: 10.1111/j.1369-7625.2007.00441.x] [Medline: 17678513]
23. Blanchard CG, Labrecque MS, Ruckdeschel JC, Blanchard EB. Information and decision-making preferences of hospitalized adult cancer patients. Soc Sci Med 1988;27(11):1139-1145. [doi: 10.1016/0277-9536(88)90343-7] [Medline: 3206248]
24. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]
25. Beisecker AE. Aging and the desire for information and input in medical decisions: patient consumerism in medical encounters. Gerontologist 1988 Jun;28(3):330-335. [doi: 10.1093/geront/28.3.330] [Medline: 3396915]
26. Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision making: a narrative review. Patient Educ Couns 2006 Feb;60(2):102-114. [doi: 10.1016/j.pec.2005.02.003] [Medline: 16442453]
27. Murray E, Pollack L, White M, Lo B. Clinical decision-making: Patients' preferences and experiences. Patient Educ Couns 2007 Feb;65(2):189-196. [doi: 10.1016/j.pec.2006.07.007] [Medline: 16956742]
28. Noncommunicable diseases country profiles 2018. World Health Organization. 2018. URL: https://www.who.int/nmh/publications/ncd-profiles-2018/en/ [accessed 2020-11-05]
29. Schachner T, Gross C, Hasl A, V Wangenheim F, Kowatsch T. Deliberative and Paternalistic Interaction Styles for Conversational Agents in Digital Health: Procedure and Validation Through a Web-Based Experiment. J Med Internet Res 2021 Jan 29;23(1):e22919 [FREE Full text] [doi: 10.2196/22919] [Medline: 33512328]
30. Zhang Z, Mai Y, Yang M. Basic and Advanced Statistical Power Analysis. Package 'WebPower'. 2018. URL: https://cran.r-project.org/web/packages/WebPower/WebPower.pdf [accessed 2020-09-21]
31. Borenstein M, Cohen J, Rothstein HR, Pollack S, Kane JM. Statistical power analysis for one-way analysis of variance: A computer program. Behavior Research Methods, Instruments, & Computers 1990 May 1;22(3):271-282. [doi: 10.3758/BF03209816]
32. White R, Walker P, Roberts S, Kalisky S, White P. Bristol COPD Knowledge Questionnaire (BCKQ): testing what we teach patients about COPD. Chron Respir Dis 2006;3(3):123-131 [FREE Full text] [doi: 10.1191/1479972306cd117oa] [Medline: 16916006]
33. Bitly. URL: https://bitly.com/ [accessed 2021-05-17]
34. Global Initiative for Chronic Obstructive Lung Disease. Pocket Guide to COPD Diagnosis, Management and Prevention: A Guide for Health Care Professionals. 2018. URL: https://goldcopd.org/wp-content/uploads/2018/02/WMS-GOLD-2018-Feb-Final-to-print-v2.pdf [accessed 2021-05-17]

XSL•FO
RenderX

35. Bildungssystem. Bundesamt für Statistik. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungssystem.html [accessed 2020-09-15]
36. EDUCATION INDICATORS: An International Perspective. National Center for Educational Statistics. URL: https://nces.ed.gov/pubs/eiip/eiip1s01.asp [accessed 2020-10-01]
37. International Standard Classification of Education: ISCED 2011. UNESCO Institute for Statistics. 2011. URL: http://uis.unesco.org/sites/default/files/documents/isced-2011-en.pdf [accessed 2020-11-11]
38. van Griethuijsen RALF, van Eijck MW, Haste H, den Brok PJ, Skinner NC, Mansour N, et al. Global Patterns in Students' Views of Science and Interest in Science. Res Sci Educ 2014 Nov 9;45(4):581-603. [doi: 10.1007/s11165-014-9438-6]
39. Komiak SY, Benbasat I. The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. MIS Quarterly 2006;30(4):941. [doi: 10.2307/25148760]
40. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
41. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: The working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
42. Lee S, Choi J. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies 2017 Jul;103:95-105. [doi: 10.1016/j.ijhcs.2017.02.005]
43. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychotherapy Research 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
44. Frey U, Suki B. Complexity of chronic asthma and chronic obstructive pulmonary disease: implications for risk assessment, and disease progression and control. The Lancet 2008 Sep 20;372(9643):1088-1099 [FREE Full text] [doi: 10.1016/S0140-6736(08)61450-6] [Medline: 18805337]
45. Van Dulmen AM, Bensing JM. Health promoting effects of the physician-patient encounter. Psychology, Health & Medicine 2010 Aug 19;7(3):289-300. [doi: 10.1080/13548500220139421]
46. Ma T, Sharifi H, Chattopadhyay D. Virtual Humans in Health-Related Interventions: A Meta-Analysis. In: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 2, 2019; Glasgow, Scotland p. 1-6. [doi: 10.1145/3290607.3312853]
47. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the Al-chatbot approach. In: PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 2017; Barcelona, Spain p. 261-265. [doi: 10.1145/3154862.3154914]
48. Gibson PG, Talbot PI, Toneguzzi RC. Self-management, autonomy, and quality of life in asthma. Population Medicine Group 91C. Chest 1995 Apr;107(4):1003-1008. [doi: 10.1378/chest.107.4.1003] [Medline: 7705105]
49. 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]
50. Montori VM, Gafni A, Charles C. A shared treatment decision-making approach between patients with chronic conditions and their clinicians: the case of diabetes. Health Expect 2006 Mar;9(1):25-36 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00359.x] [Medline: 16436159]
51. Ousseine YM, Durand M, Bouhnik A, Smith A, Mancini J. Multiple health literacy dimensions are associated with physicians' efforts to achieve shared decision-making. Patient Educ Couns 2019 Nov;102(11):1949-1956. [doi: 10.1016/j.pec.2019.05.015] [Medline: 31130338]
52. Reach G. Patient autonomy in chronic care: solving a paradox. Patient Prefer Adherence 2013 Dec 12;8:15-24 [FREE Full text] [doi: 10.2147/PPA.S55022] [Medline: 24376345]
53. Bérubé C, Schachner T, Keller R, Fleisch E, V Wangenheim F, Barata F, et al. Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review. J Med Internet Res 2021 Mar 29;23(3):e25933 [FREE Full text] [doi: 10.2196/25933] [Medline: 33658174]

Abbreviations

AI: artificial intelligence
BCKQ: Bristol COPD Knowledge Questionnaire
CA: conversational agent
COPD: chronic obstructive pulmonary disease
GOLD: Global Initiative for Chronic Obstructive Lung Disease
ISCED: International Standard Classification of Education

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.12.20; peer-reviewed by E Dogan, D Dhinagaran, R Milani, E Bellei; comments to author 15.01.21; revised version received 12.02.21; accepted 14.04.21; published 26.05.21.

Please cite as:
Gross C, Schachner T, Hasl A, Kohlbrenner D, Clarenbach CF, Wangenheim FV, Kowatsch T
Personalization of Conversational Agent-Patient Interaction Styles for Chronic Disease Management: Two Consecutive Cross-sectional Questionnaire Studies
J Med Internet Res 2021;23(5):e26643
URL: https://www.jmir.org/2021/5/e26643
doi:10.2196/26643
PMID:33913814

©Christoph Gross, Theresa Schachner, Andrea Hasl, Dario Kohlbrenner, Christian F Clarenbach, Forian V Wangenheim, Tobias Kowatsch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2021. 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

Priority Given to Technology in Government-Based Mental Health and Addictions Vision and Strategy Documents: Systematic Policy Review

Shalini Lal[1,2,3], BScOT, MSc, PhD; Lyna Siafa[2,4]; Hajin Lee[1,2], PhD; Carol E Adair[5,6], PhD

[1]School of Rehabilitation, Faculty of Medicine, University of Montréal, Montréal, QC, Canada

[2]Youth Mental Health and Technology Lab, Health Innovation and Evaluation Hub, University of Montréal Hospital Research Centre, Montréal, QC, Canada

[3]Prevention and Early Intervention Program for Psychosis and ACCESS Open Minds, Douglas Mental Health University Institute, Montréal, QC, Canada

[4]Department of Medicine, McGill University, Montréal, QC, Canada

[5]Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[6]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Shalini Lal, BScOT, MSc, PhD
School of Rehabilitation
Faculty of Medicine
University of Montréal
PO Box 6128, Centre-ville station
Montréal, QC, H3C 3J7
Canada
Phone: 1 514 890 8000 ext 31581
Email: shalini.lal@umontreal.ca

Abstract

Background: The use of information and communication technologies (ICTs) to deliver mental health and addictions (MHA) services is a global priority, especially considering the urgent shift towards virtual delivery of care in response to the COVID-19 pandemic. It is important to monitor the evolving role of technology in MHA services. Given that MHA policy documents represent the highest level of priorities for a government's vision and strategy for mental health care, one starting point is to measure the frequency with which technology is mentioned and the terms used to describe its use in MHA policy documents (before, during, and after COVID-19). Yet, to our knowledge, no such review of the extent to which ICTs are referred to in Canadian MHA policy documents exists to date.

Objective: The objective of this systematic policy review was to examine the extent to which technology is addressed in Canadian government–based MHA policy documents prior to the COVID-19 pandemic to establish a baseline for documenting change.

Methods: We reviewed 22 government-based MHA policy documents, published between 2011 and 2019 by 13 Canadian provinces and territories. We conducted content analysis to synthesize the policy priorities addressed in these documents into key themes, and then systematically searched for and tabulated the use of 39 technology-related keywords (in English and French) to describe and compare jurisdictions.

Results: Technology was addressed in every document, however, to a varying degree. Of the 39 searched keywords, we identified 22 categories of keywords pertaining to the use of technology to deliver MHA services and information. The 6 most common categories were *tele* (n=16/22), *phone* (n=12/22), *tech* (n=11/22), *online* (n=10/22), *line* (n=10/22), and *web* (n=10/22), with n being the number of policy documents in which the category was mentioned out of 22 documents. The use of terms referring to advanced technologies, such as *virtual* (n=6/22) and *app* (n= 4/22), were less frequent. Additionally, policy documents from some provinces and territories (eg, Alberta and Newfoundland and Labrador) mentioned a diverse range of ICTs, whereas others described only 1 form of ICT.

Conclusions: This review indicates that technology has been given limited strategic attention in Canadian MHA policy. Policy makers may have limited knowledge on the evidence and potential of using technology in this field, highlighting the value for knowledge translation and collaborative initiatives among policy makers and researchers. The development of a pan-Canadian

XSL•FO
RenderX

framework for action addressing the integration and coordination of technology in mental health services can also guide initiatives in this field. Our findings provide a prepandemic baseline and replicable methods to monitor how the use of technology-supported services and innovations emerge relative to other priorities in MHA policy during and after the COVID-19 pandemic.

(J Med Internet Res 2021;23(5):e25547) doi:10.2196/25547

KEYWORDS

e-mental health; digital mental health; virtual care; technology; mental health; addictions; review; policy; e-health; mHealth; digital health; tele

Introduction

According to a Canadian survey conducted in 2018 [1], approximately 5.3 million individuals aged 12 years and older reported a need for mental health care in the previous year, yet the mental health needs of almost half of these individuals were unmet. The need to improve access and quality of mental health and addictions (MHA) services is increasingly reflected in Canadian mental health policy and practice [2]. Moreover, consequences of COVID-19 have compounded the need for mental health services [3,4], with concurrent public health guidelines for social distancing creating additional challenges to the delivery of care. These challenges have driven rapid integration of technology into the MHA system [5-8].

The use of the internet and related information and communication technologies (ICTs) to deliver or enhance mental health services (also referred to as digital mental health/digital health, e-mental health/e-health, telehealth, or virtual care) is well recognized for its potential in helping to overcome some of the barriers that individuals face in accessing mental health care [9-12]. For example, compared to conventional approaches, the use of ICTs can provide more accessible, empowering, and sustainable care [11-14]; improve access to specialist mental health care; and reduce wait times [15]. This is particularly important for rural and remote areas that have inequitable access to mental health services due to geographic distance and scarcity of resources, and for populations that are less inclined to seek help in person due to stigma [11,12]. Moreover, opportunities to receive care via ICTs also have the potential to provide individuals choice and control over their care [16].

A wide range of ICTs (eg, telephone and videoconferencing services, websites, smartphone apps, social media) have been used to develop innovations that deliver mental health services [11,12]. In general, these innovations serve the purpose of providing information and self-management tools; conducting screening and assessment; facilitating monitoring of symptoms, activities, and behaviors; and delivering psychological and social interventions (eg, peer support, counseling, case management) [11,12]. Examples include telephone-based crisis interventions [17], psychiatric assessment and treatment via videoconferencing (also known as telepsychiatry) [18,19], web-based therapeutic interventions [20,21], behavioral and psychosocial treatments using smartphone software apps and secured websites [22,23], use of sensors for patient monitoring [24], online peer support through social media [25], and virtual reality therapies [26].

In terms of evidence, there is increasing research indicating the efficacy of using ICTs to deliver mental health services and its effectiveness [20,27,28]. For example, a recent review and meta-analysis demonstrated the efficacy of using ICT-based interventions to address eating disorders under controlled trials. Specifically, digital interventions were shown to be more effective than control conditions in reducing risk factors and symptoms in prevention-focused trials and treatment-focused trials [28]. However, other types of individual-level research, including economic analyses, are still needed, in addition to broader analyses, such as studying change at the policy level.

In Canada, innovation, research, and practice on the use of ICTs in mental health care has significantly advanced over the past decade. Several examples of such advancements are provided in the gray and peer-reviewed literature [10,11], including but not exclusive to evidence-based technology-supported initiatives from the Strongest Families Institute (a not-for-profit organization originating from clinical trials conducted at the IWK Health Centre (formerly Izaak Walton Killam Health Centre) in Nova Scotia and onlinetherapyuser.ca (based on clinical research conducted at the University of Regina with eventual sustained funding received from the Saskatchewan Ministry of Health). However, the use of ICTs in mental health care is still a relatively new paradigm [11,29,30], with implementation remaining limited and fragmented [11,29] even within the context of the COVID-19 pandemic. This indicates a systemic gap between evidence and its implementation, which may in part be due to limited guidance on how to translate evidence into policy and competing priorities in research and government contexts [31,32]. More research at the policy level can inform and bridge the evidence-policy-practice gap [33] and can inform effective collaboration on the development of policy among academics, researchers and developers, health care planners and policy makers, practitioners, and consumer representatives [11,29,32].

Mental health policy and action documents have a critical role and responsibility for optimizing technology integration within the mental health care system. The impact of such a role is well illustrated by the policy example of Australia, a country that is at the forefront of integrating technology in the delivery of mental health services. In 2012, the Australian government's Department of Health and Aging produced a policy document entitled, "E-Mental Health Strategy for Australia," which provided a vision for e-mental health services and key areas for action. The strategy also provided guidance on investment and development to advance the field and was created through input from expert advisors representing top researchers within the field, consumer representatives, and executive leads of organizations delivering mental health services using technology [34].

XSL•FO
RenderX

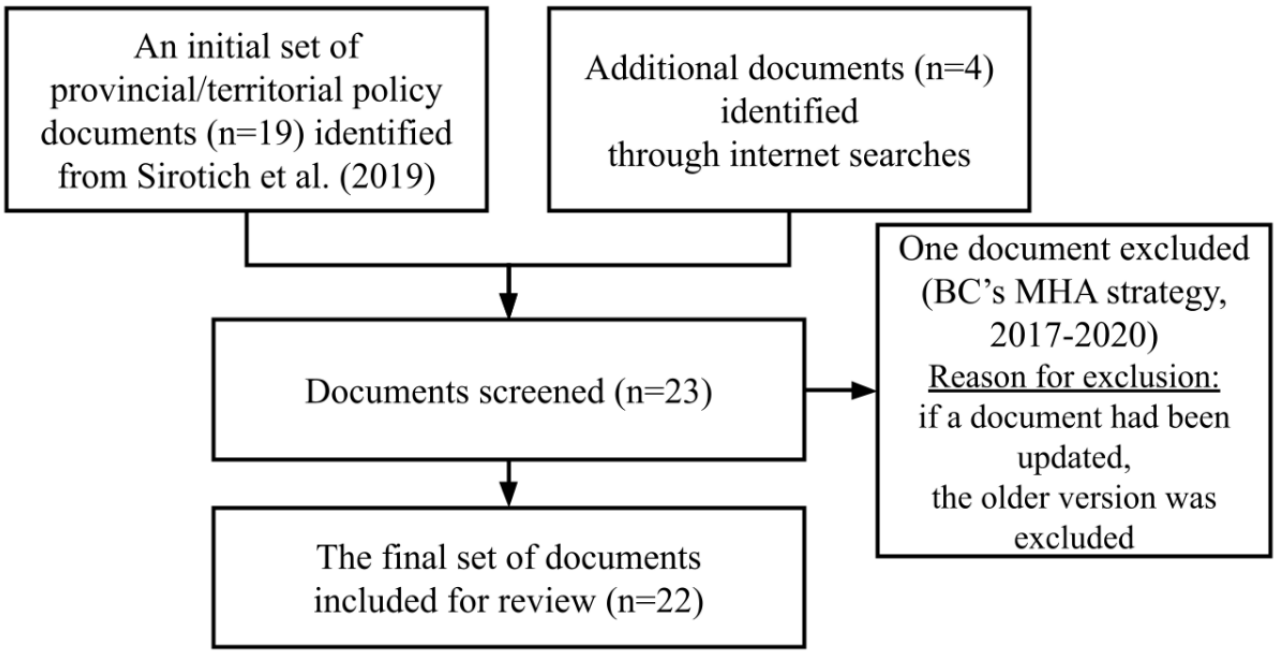
The role of technology in mental health policy is even more critical considering the COVID-19 pandemic. Given this, it is important to understand the priority given to technology in government policy and to monitor the advancement of this priority at the policy level before, during, and after the pandemic. In 2017, a systematic review was published with the aim of comparing areas of focus on the use of technology in pediatric mental health care by examining research studies, and government and organizational documents. It included 2 Canadian-based policy documents, 1 at the national level, and 1 from Ontario [32]. However, to our knowledge, no systematic review has focused on the extent to which technology is discussed in mental health policies from all Canadian provinces and territories to date. Thus, the objective of this systematic policy review is to determine the extent to which technology is considered in Canadian-based mental health policy documents. Specifically, we aimed to synthesize the policy priorities addressed in government-based MHA documents and to better understand the extent to which technology is described as having a role in achieving these policy priorities.

Methods

Collection and Selection of MHA Policy Documents

As illustrated in Figure 1, we started with an initial set of 19 provincial/territorial (P/T) policy documents published until June 2017 that were identified as part of a project led by CEA on the development of a pan-Canadian MHA performance measurement framework [2,35]. These initial documents were obtained from a systematic internet search of P/T government websites. They were then confirmed for completeness for each jurisdiction in consultation with researchers, data managers, and policy makers from the Mental Health Commission of Canada's (MHCC) Provincial and Territorial Advisory Group and through consensus by the research team based on inclusion criteria pertaining to policy timeframes, focus on special populations, and versions of documents [2,35]. Two additional team members from this current review (SL and LS) considered each of these 19 documents to determine if they met our inclusion criterion for the current review and to assess the need for additional searches. The inclusion criterion was a document that describes a P/T vision or strategy for mental health service delivery; we also endeavored to ensure that each P/T was represented by a mental health policy document. In the exceptional case of Nunavut, which became the newest Canadian territory in 1999 at which time a vision or strategy document did not exist, the next closest type of document was included [36]. The exclusion criterion was any older document for which an updated version of the same type of document from that P/T was available. This was to provide the most updated perspective of mental health policy across the P/T at the time of conducting the review.

Figure 1. Identification and selection of MHA policy documents. MHA: mental health and addictions.



To identify updated versions of the policy documents, during the summer of 2019, we conducted internet searches (on the P/T websites and using the Google search engine). Based on this search, we identified 4 additional documents: (1) a recent MHA policy document from British Columbia (2019-2029) [37], (2) Nova Scotia's progress update on the 5-year plan (2016) [38], (3) an action plan for suicide prevention in Nunavut (2016-2017) [39], and (4) the Final Report of Ontario's MHA Leadership Advisory Council (2017) [40]. Thus, we had a total set of 23 documents, from which we excluded 1, BC's MHA Strategy (2017-2020) [41] because a newer version was available; this resulted in a final set of 22 documents. Each of these documents is presented in Multimedia Appendix 1, including title, source in terms of P/T, URL links, and overview.

All documents described a vision or framework, strategy or action plan, or progress update related to MHA policy in that province/territory, 20 documents addressed MHA policies

XSL•FO
RenderX

pertaining to the P/T's population as a whole, while 2 documents were about MHA policy specific to a subpopulation in that P/T (ie, the 2013 BC First Nations and Aboriginal People's MHA 10 Year Plan and the 2014 Yukon's Child and Youth MHA Framework (2014) [42,43].

Data Extraction and Analysis

We first reviewed the documents using an initial list of 9 technology words (*e-mental health*, *tech*, *tele*, *net*, *internet*, *web*, *virtual*, *app*, and *digital*) to gain a preliminary impression of how technology was considered in the government documents. These 9 words were based on SL and CEA's previous review of the literature on technology in mental health care [12] and recent academic and gray literature (eg, e-mental health documentation from the MHCC [10]). We reviewed the documents systematically using the control-F function and recorded instances where the 9 words were mentioned in a Microsoft Word document table. We did not find many examples of policy documents mentioning the use of technology; however, we felt that perhaps this could have been due to the limited number of words we used to search the documents. The initial search process informed our understanding of the types of terms used in the policy documents and led to the development of a more detailed protocol to guide our review. Our protocol included data extraction tables for organizing the data (eg, name of policy document, year of publication, source, excerpts from the documents illustrating use of a technology term). As part of this protocol, we also created a process for validating the data extraction and analysis to enhance reliability of our results. Our final protocol included 39 technology-related keywords that would be used to examine the extent to which each policy document addressed the topic of technology in relation to the delivery of mental health services. This list was finalized through pilot testing in several of the documents before being implemented systematically across all the documents. We also translated this keywords list into French, first by a native French-speaking team member, with initial validation by a second native French-speaking team member (LS) and final validation by the project lead (SL) who is a native English-speaking team member. The keywords list in English and French is presented in Table 1.

Next, we reviewed each policy document to gain a better understanding of the type of document (ie, vision, strategy, and progress related to mental health policy). We determined the type of document mainly based on their titles, for example, documents that typically used the terms *vision*, *strategy*, or *update* in the titles. We also conducted a modified conventional content analysis [44] of the MHA policy priorities described in these documents to obtain a snapshot of the recurrent themes addressed across the set of 22 documents. This involved reviewing the content of each document with a focus on extracting the MHA priorities using a data extraction table in a Microsoft Excel file (LS). The priorities were then coded inductively (ie, staying close to the terms used) and then grouped to arrive at a list of 14 categories. These categories were then reviewed by 2 members of the team (SL and LS) to discuss overlaps, redundancies, and what each category represented in terms of content. This led to a final coding framework of 10 MHA priorities that were represented across the 22 documents. Finally, each of the documents was then re-examined by LS to tabulate which of the 10 MHA priorities was addressed by each of the P/Ts, and this tabulation was then validated by HL.

Next, using the control-F function, we searched each of the documents using the keyword terms listed in Table 1. All findings that represented the use of technology within the context of mental health care were charted in a Microsoft Excel file. The data in this Excel file were initially extracted by a member of our team (LS) and validated by another (HL).

After we completed the search for the keywords and all instances of the use of technology were extracted from the policy documents, we grouped the 39 keywords into 22 categories to facilitate the synthesis process. Specifically, we merged the synonyms and abbreviations of a keyword into the same category. For example, *electronic* and the abbreviation *EHR* (*electronic health record*) were grouped into the *electronic* category. We also removed keywords that had not been mentioned in any of the policy documents. For example, *EMR* (*electronic medical record*) was removed, since it was not mentioned in any document. This final set of 22 categories was arrived at through an iterative process involving discussion and consensus between SL and LS. We then organized our findings based on the number of policy documents (n) that mentioned each category out of the 22 policy documents (eg, the keyword category *Tele* was mentioned in 16 out of the 22 policy documents; n=16/22; Table 2); subsequently, we identified which P/Ts had documents mentioning each category (Table 2). We also counted the number of categories mentioned by each P/T out of the total 22 categories, as represented by n (eg, only 1 keyword category, *electronic*, was mentioned by the Northwest Territories; Multimedia Appendix 3). Note that we counted compound keywords twice to maintain a systematic approach to the keyword search of the policy documents. For example, *telephone* was counted once in *tele* and once in *phone*, and *text messages* was counted once in *text* and once in *messag**.

XSL•FO
RenderX

Table 1. List of English and French keywords[a] used to search the documents.

Keywords list (English)	Keywords list (French)
app (eg, application)	appli (eg, application)
artificial intelligence	intelligence artificielle
avatar	avatar
bots (eg, chatbots, chat bots, robots), chat (eg, live chat, live-chat)	bots (eg, robots), chat (eg, chatbots, chatter, chat en ligne, chat en direct), dialogueur, agent conversationnel
messag*, instant messaging	messagerie instantanée
cell, phone, line, link	cellulaire, téléphone
mobile (eg, mobile device, mobile health), mhealth	mobile (eg, appareil mobile, santé mobile) msanté (eg, m-Santé, m-santé)
text, SMS	message texte, texto, SMS (eg, message-texte)
smart (eg, smartphone, smart phone, smart watch, smartwatch)	intelligent (eg, téléphone intelligent, montre intelligente)
computer (eg, laptop computer, desktop computer, personal computer), laptop	ordinateur, portable (eg, ordinateur portable, ordinateur-portable)
cyber (eg, cyber-psychology, cyber psychology, cyberspace, cyber-space)	cyber (eg, cyberpsychologie, cyber-psychologie, cyber psychologie, cyber-espace, cyber espace)
device (eg, wearable device, wearable-device, mobile device, portable device)	appareil (eg, appareil portable)
digital	digital
e-mental health	santé mentale électronique, e-santé mentale, santé numérique, cybersanté mentale
email	courriel, courrier électronique
electronic, EMR[b], EHR[c], e-referral (eg, mail, electronic-mail, electronic case management, electronic-case-management, electronic medical record, electronic-medical-record, e-patient, e patient, electronic patient, electronic-patient, e-referral, e referral)	électronique, DME[d] (eg, gestion de cas électronique, dossier médical électronique, patient électronique, référence électronique, référencement électronique)
net (eg, internet), web (eg, website)	net (eg, internet), toile, web (eg, site web)
online	en ligne
portals	portail
sensor	senseur
platform, (Headspace), (ACCESS[e] Open Minds[f])	plateforme, plate-forme
social media (eg, social-media)	médias sociaux, réseaux sociaux
tech (eg, technology, communication technology, health information and communication technology)	techno (eg, technologie, technologique, technologies de l'information et de la communication en santé)
tele (eg, telehealth, tele-health, tele-medicine, telemedicine, tele-mental-health, telementalhealth, tele-psychiatry, telepsychiatry, tele-psychology, telepsychology, tele-therapy, teletherapy)	télé (eg, télé-santé, télé santé, télé-médecine, télé médecine, télé santé mentale, télé-psychiatrie, télé psychiatrie, télé-psychologie, télé psychologie, télé-thérapie, télé thérapie)
video (eg, video-health, videohealth, video mental health, videomental-health, video-psychiatry, videopsychiatry, video-psychology, videopsychology, video-therapy, videotherapy)	vidéo (eg, vidéo-santé, vidéo santé, vidéo santé mentale, vidéo-psychiatrie, vidéo psychiatrie, vidéo-psychologie, vidéo psychologie, vidéo-thérapie, vidéo thérapie)
virtual (eg, virtual reality, virtual-reality), VR[g]	virtuel (eg, réalité virtuelle)

[a]The original list of 9 words included *e-mental health*, *tech*, *tele*, *net*, *internet*, *web*, *virtual*, *app*, and *digital*.

[b]EMR: electronic medical record.

[c]EHR: electronic health record.

[d]DME: dossier médical électronique.

[e]ACCESS: Adolescent/young adult, Connecting to, Community-driven, Early, Strengths-based and Stigma-free services.

[f]A national network connecting youth, families, caregivers, researchers, service providers, and policy makers to improve youth mental health care across Canada.

[g]VR: virtual reality.

XSL•FO
RenderX

Table 2. Frequency of technology-related categories mentioned in the provincial/territorial mental health and addictions policy documents.

Category	Examples of keywords[a]	Policy documents mentioning the category[b], n	Province/territory mentioning the category (item #)[c]
Tele (*télé* in French)	telehealth, TeleHealth, tele-health, telephone, telepsychiatry, tele-psychiatry, telephone counselling, teleconferencing, telecounselling, telehealth network, telehealth consultation, video teleconferencing, tele-mental health, telemedicine, *consultation téléphonique psychosociale*	16	Alberta (2), British Columbia (3 and 4), Manitoba (5), Newfoundland and Labrador (7 and 8), Nova Scotia (9 and 10), Nunavut (13), Ontario (14 and 17), Prince Edward Island (18), Québec (19), Saskatchewan (20), Yukon (21 and 22)
Phone (*téléphone* in French)	phone, Kids Help Phone, telephone, telephone counselling, telephone screening, crisis phone lines, phone line, *téléphone*, phone-based coaching service	12	Alberta (1), Newfoundland and Labrador (7 and 8), Nova Scotia (9 and 10), Nunavut (13), Ontario (14), Prince Edward Island (18), Québec (19), Saskatchewan (20), British Columbia (4), Yukon (21)
tech/techno	technology-based services, technology-based solutions, technologies, information technology, communication technology, technology, e-health technology, technology-based interventions, communication technologies, technological solutions, distance technology, emerging technology, evolving technologies, *technologie*	11	Alberta (1 and 2), Manitoba (5), New Brunswick (6), Newfoundland and Labrador (7 and 8), Ontario (16 and 17), Québec (19), Saskatchewan (20), Yukon (22)
online	online counselling, online, online information, online cognitive behavioural therapy, online self-help supports, online training, online program, online options, online delivery, on-line resource, healthlineonline.ca, online gateways, provincial online hub, centralized online hub, online psychotherapy, ReachOut.com online platform, 211 online directory of community services (sk.211.ca), online and distance supports, online knowledge exchange (KE)[d] space, online access point, online clinical treatments, online treatments, online resources, online directory, online tools, online modules, online innovation hub, online prevention, online clinical supports, online application, online chat capabilities, online experience, online government resources, online access, online peer support tools	10	Alberta (1), British Columbia (4), Newfoundland and Labrador (7 and 8), Nova Scotia (10), Nunavut (13), Ontario (14 and 15), Saskatchewan (20), Yukon (22)
line	distress lines, Mental Health Crisis Line, helpline, Provincial Warm Line, crisis line, Problem Gambling Helpline, Kamatsiaqtut Help Line, crisis phone lines, HealthLine (811), healthlineonline.ca, phone line, community-based crisis lines, Kids Help Line, crisis helpline, line, crisis lines network	10	Alberta (1), British Columbia (4), Newfoundland and Labrador (7 and 8), Nova Scotia (9 and 10), Nunavut (13), Saskatchewan (20), Yukon (21 and 22)
web	web-based, website, mental health website, youth mental health websites, web resource, web application	10	Alberta (1 and 2), British Columbia (4), Manitoba (5), New Brunswick (6), Newfoundland and Labrador (7 and 8), Ontario (14), Saskatchewan (20), Yukon (21)
virtual	virtual resources, virtual counselling, virtual solutions, virtually, virtual reality, virtual teams, virtual mental health counselling	6	Alberta (1 and 2), British Columbia (4), Newfoundland and Labrador (7 and 8), Yukon (22)
electronic	electronic health record, EHR[e], electronically, electronic client record, electronic mental health information and support, electronic information and support, electronic links	6	Alberta (1), Newfoundland and Labrador (7), Northwest Territories (11), Ontario (14), Prince Edward Island (18), Yukon (21)
social media	social media, social media communications	5	British Columbia (3), Newfoundland and Labrador (7 and 8), Nunavut (13), Ontario (15)
video	video conference, videos, videoconferencing, video teleconferencing	4	Newfoundland and Labrador (7), Nova Scotia (9), Ontario (14), Saskatchewan (20)
app	mobile applications, app, application, online application, web application	4	Alberta (2), Nova Scotia (9), Yukon (21 and 22)

XSL•FO
RenderX

Category	Examples of keywords[a]	Policy documents mentioning the category[b], n	Province/territory mentioning the category (item #)[c]
link	Health Link, links, electronic links	4	Alberta (1 and 2), Ontario (14), Yukon (21)
text	text, text messages	3	Newfoundland and Labrador (7 and 8), Nova Scotia (10)
e-mental health	e-mental health, e-mental health services, e-Mental Health and Addictions platform (eMHA[f])	3	Newfoundland and Labrador (7 and 8), Saskatchewan (20)
platform	platform, MyHealth platform, mental health and addictions platform, e-Mental Health and Addictions platform (eMHA), central platform, integrated eMHA platform, online platform	3	Alberta (1 and 2), Saskatchewan (20)
net	internet, internet cognitive behavioural therapy, iCBT[g]	3	British Columbia (4), Newfoundland and Labrador (7), Saskatchewan (20)
chat	chat, online chat capabilities, web-based chat capabilities	3	British Columbia (4), Nova Scotia (10), Yukon (21)
digital	digital therapy, digital view	2	Alberta (1), Newfoundland and Labrador (7)
portal	portal, web-based portal	2	British Columbia (4), Ontario (14)
mobile	mobile applications	1	Alberta (2)
messag*	text messages	1	Nova Scotia (10)
e-referral	e-referral tool	1	Alberta (1)

[a]Examples of keywords for each category were systematically extracted from the policy documents reviewed.

[b]Compound keywords were counted twice to maintain a systematic approach to the keyword search of the policy documents. For example, *telephone* was counted once in *tele* and once in *phone*, and *text messages* was counted once in *text* and once in *messag**.

[c]The item number in parentheses reflects the item number in Multimedia Appendix 1.

[d]KE: knowledge exchange.

[e]EHR: electronic health record.

[f]eMHA: e-Mental Health and Addictions platform.

[g]iCBT: internet cognitive behavioural therapy.

Results

Description of the MHA Policy Documents

Most Canadian P/T governments (including Manitoba, New Brunswick, Northwest Territories, Prince Edward Island, Québec, and Saskatchewan) had 1 MHA policy document [45-50]. Alberta [51,52], British Columbia [37,42], Newfoundland and Labrador [53,54], Nova Scotia [38,55], Ontario [40,56-58], and Yukon [43,59] had two or more MHA policy documents. Nunavut did not have any MHA policy documents; instead, a statement of overall government directions and an action plan for suicide prevention in Nunavut were used in the analysis [36,39]. The types of policy documents we reviewed were visions or frameworks, strategies or action plans, and progress updates. The majority (n=17/22) of the policy documents were focused on MHA policy strategies and action plans. The remaining 5 documents were about visions or frameworks for an MHA system (n=2/22), progress updates on the MHA strategies or action plans (n=1/22), or combined MHA strategies or action plans and progress updates (n=2/22). We extracted data on the overview of each MHA policy document, particularly focusing on P/T policy priorities addressed in the documents (see Multimedia Appendix 1).

Our synthesis of P/T MHA policy priorities identified 10 themes, with 7 themes being found in the policies of all 13 P/Ts: wellness/recovery/MHA promotion, prevention, and early intervention (eg, promoting good mental health, reducing stigma); service integration (eg, integration of service delivery to improve the currently fragmented system); collaboration (eg, promoting a multidisciplinary and collaborative team approach); children, youth, and families (eg, youth suicide prevention, strongest families program); improving access (eg, better access to services for remote, rural, and underserved communities; decreasing wait times and travel times); cultural safety and indigenous communities (eg, culturally relevant treatments and services, health equity for diverse populations); and use of technology (eg, sharing of evidence-based MHA information or services online). The remaining 3 themes were support for seniors (n=12/13); innovation, improvement, and research (n=2/13); and reducing bullying (n=6/13). As depicted in Multimedia Appendix 2, we organized the policy priorities as themes related to a population or a service; 3 of the priorities related to specific populations (children, youth, and seniors; and diversity of communities, such as indigenous communities, racialized groups, new arrivals, and persons with disability), and the remaining 7 were related to a service priority.

XSL•FO
RenderX

Use of Technology in the MHA Policy Documents

Of the 39 keywords listed in Table 1, we identified 22 categories addressing the use of technology to deliver mental health services listed in Table 2. Table 2 also includes examples of keywords from each category that were used in the policy documents.

The 6 most common keyword categories were *tele* (n=16/22), *phone* (n=12/22), *tech* (n=11/22), *online* (n=10/22), *line* (n=10/22), and *web* (n=10/22). Among the 6 most common categories, 3 pertained to connecting people remotely in real-time: *tele* (eg, *telephone*, *telehealth*, *telepsychiatry*; n=16/22), *phone* (eg, *telephone counselling*, *crisis phone lines*, *phone-based coaching service;* n=12/22), and *line* (eg, *distress lines*, *mental health crisis line*, *helpline*; n=10/22). Policy documents also referred to *technology-based services* in general without detailing specific technologies. As such, *tech* (n=11/22) was the third most common category, followed by *online* (n=10/22), *line* (n=10/22), and *web* (n=10/22). Examples of words mentioned in this regard include *technology-based services*, *technology-based solutions*, and *technologies*; *online counselling*, *online information*, and *online cognitive behavioural therapy*; and *web-based*, *website*, and *mental health website*.

In this review, *high-tech* was used to describe technologies other than telephone-based services (eg, live chat on a website, the use of smartphone apps) and *low-tech* was used to describe telephone-based services (eg, crisis telephone help lines). High-tech categories were less common and included *virtual* (n=6/22), *electronic* (n=6/22), *social media* (n=5/22), *video* (n=4/22), *app* (n=4/22), and *link* (n=4/22). Examples of words mentioned in this regard include *virtual resources*, *virtual counselling*, and *virtual reality*; *electronic health record* and *electronically*; *social media* and *social media communications*; *video conference* and *videos*; *mobile applications*, *app*, and *online application*; and *Health Link*, *links*, and *electronic links*.

Other high-tech categories and broader approaches to the use of technology were mentioned at an even lower frequency: *text* (eg, *text*, *text messages*), *e-mental health* (eg, e-*mental health*, e-*mental health services*, e-*Mental Health and Addictions platform*), *platform* (eg, *platform*, *MyHealth platform*, *mental health and addictions platform*, *online platform*), *net* (eg, *internet*, *internet cognitive behavioural therapy*, [*iCBT*]), and *chat* (eg, *chat*, *online chat capabilities*, *web-based chat capabilities*) were each mentioned in 3 policy documents (n=3/22). *Digital* (eg, *digital therapy*, *digital view*) and *portal* (eg, *portal*, *web-based portal*) were each mentioned in 2 policy documents (n=2/22). *Mobile* (eg, *mobile applications*), *messag** (eg, *text messages*), and *e-referral* (eg, *e-referral tool*) were each mentioned in 1 policy document (n=1/22). Overall, there was a diverse range of high-tech categories mentioned in the P/T policy documents but these were mentioned at a much lower frequency compared to the low-tech categories.

Additionally, as shown in Multimedia Appendix 3, there was a high level of variation in the extent to which the policy documents referred to the 22 categories across the 13 P/Ts of Canada: Alberta (n=14/22), Newfoundland and Labrador (n=14/22), Yukon (n=11/22), British Columbia (n=10/22), Ontario (n=10/22), Saskatchewan (n=10/22), Nova Scotia (n=9/22), Nunavut (n=5/22), Manitoba (n=3/22), Prince Edward Island (n=3/22), Québec (n=3/22), New Brunswick (n=2/22), and Northwest Territories (n=1/22).

As noted above, more than half of the technology categories were mentioned in the policy documents of Alberta and Newfoundland and Labrador. Alberta's MHA action plans (2015, 2017-2020) emphasize developing and implementing a range of technology-based solutions (eg, websites, telehealth, mobile apps, virtual solutions) for MHA problems, and combining and enhancing resources that are already available (ie, telephone-based interventions such as Health Link and Kids Help Phone) [51,52]. For example, in its 2017-2020 MHA action plan, it states the following: "develop virtual, technology-based solutions to help people access tools, information and treatment to address addiction and mental health issues," and "share information on websites, telehealth, mobile, mobile applications, and other technologies" [51]. Similarly, Newfoundland and Labrador's vision for a renewed MHA system (2017) and MHA action plan (2017-2022) highlighted using a range of technologies (eg, including telehealth, videoconferencing, telephone crisis lines, online, text, virtual reality, social media, electronic health records, e-mental health) to achieve a full continuum of care and improve access to services [53,54].

Regarding the policy documents that mentioned only a few technology categories (eg, Québec, New Brunswick, and the Northwest Territories), Quebec's action plan (2015-2020) mentioned developing new technologies to offer a network of youth-friendly MHA services, reiterated the availability of 24-7 psychosocial telephone consultations, and mentioned 2 institutions that aimed to promote mental health and new technologies: the Centre national d'excellence en santé mentale (CNESM) and the Service d'accueil, d'analyse, d'orientation et de référence des services sociaux généraux (AAOR) offering services over the phone [49]. In the New Brunswick Family Plan (2017), the only reference to technology was "information related to MHA treatment options are available to the public through various means, including web-based and print materials" [46]. The only mention of technology in the Northwest Territories strategic framework (2016-2021) pertained to electronic health records to promote the integration of service delivery [47]. For more details, refer to Table 2.

Discussion

Our objective for this policy review was to determine the extent to which technology is prioritized in mental health and addictions (MHA) policy documentation. Given that policy is context-based, we focused on the Canadian context. The methods that we used in this review can help to inform future efforts to monitor the evolutionary role of technology in mental health policy in Canada and abroad. Our findings have implications for policy makers, health care leaders, and researchers working towards improving access and quality of mental health services.

XSL•FO
RenderX

Principal Findings

Our key findings are as follows. First, although all the documents referred to the use of technology for MHA care, there was a high level of variation in the extent to which technology was addressed across the P/T documents and how it was addressed. For example, Newfoundland and Labrador's MHA vision (2017) and action plan (2017-2022) described a variety of technology-based services (including telehealth, videoconferencing, telephone crisis lines, online, text, virtual reality, social media, electronic health records, e-mental health) to support access and continuum of care [53,54], whereas the strategic framework from the Northwest Territories (2016-2021) only mentioned electronic health records to promote the integration of service delivery [47], and Québec's action plan (2015-2020*)* only mentioned using technologies to offer MHA services to youth and providing psychosocial telephone consultations to the general population [49]. Second, only a few of the documents referred to the use of "high-tech" solutions, such as virtual reality and apps, with more than half mainly describing "low-tech" solutions, such as telephone-based helplines and telephone consultations for families. Considering the limitations of telephone-based interventions compared to other formats, such as video (eg, the latter resulting in fewer treatment errors and greater diagnostic and decision-making accuracy) [60] and considering mainstream society's increasing use of other formats of communication, such as texting, live chat, videoconferencing, and social media, addressing high-tech solutions in Canadian mental health policy is warranted. Third, there was inconsistency in the use of technology-related terminology across the documents, with the term *e-mental health* (the term used by the MHCC) [10,30] rarely mentioned. Fourth, there was limited strategic focus and guidance in the policy documents on how to move forward with the successful implementation of technology-enabled services within the public mental health care system.

It is important to note that the P/T policy documents reflect governmental priorities for MHA services and did not have the purpose of providing a complete perspective on the use of technology. Although it may not be the intention of the policy documents to provide an exhaustive list of initiatives, describing exemplary evidence-based initiatives (eg, onlinetherapyuser.ca) that have received government funding is warranted to provide guidance to local health authorities. Moreover, the variation that we identified across P/Ts on the degree of attention to technology in MHA care is noteworthy.

There are several implications of our findings. First, there is a need for strategic policy attention on the integration of technology in mental health; this requires addressing topics such as funding for implementation initiatives, infrastructure, regulation, capacity building, intersectoral collaboration, translation research, policy-research-practice partnerships, among others [11,12,29,32,33], as well as partnerships between researchers, practitioners, and policymakers. Advancement in these areas can be facilitated through interactions among policy makers and health services researchers specializing in the use of technology to deliver mental health care. Choi et al [61] provide several insights that can be helpful in this regard, including recommendations to incentivize collaboration among policy makers and scientists, use of knowledge brokers (to facilitate knowledge translation), and acknowledgment of the various issues affecting collaboration, including different languages and perspectives on evidence and different timeframes on achieving deliverables. Policy-making initiatives that facilitate the involvement of consumers, caregivers, and service is also important to consider.

Second, the inconsistency in the use of technology-related terms across the P/Ts is not surprising given the evolving nature of the field itself and inconsistencies in the research literature. Nonetheless, this poses challenges to conducting future policy reviews to monitor advancement in the field. To address this issue, a living document of terminology, that is updated yearly, could be made accessible to mental health policy makers to encourage a level of standardization in the language used to refer to the use of technology in mental health care.

Third, the inconsistency in the degree of attention on technology across the P/T documents may be a reflection that policy makers have limited knowledge on the evidence and potential of using technology in this field, highlighting the value for knowledge translation and collaborative initiatives among policy makers and researchers. It may also reflect the degree to which technology is prioritized at the front-line level of service delivery, raising the question of whether there is a lack of equity across the country in terms of the public's access to a wide range of services and interventions delivered through technology. This lack of attention in policy can contribute to a fragmented and unregulated approach to using technology to deliver mental health services in the country. Translational research for policy development and implementation planning is needed to better understand viewpoints from users, developers, providers, and policy makers on the implementation of technology-supported innovations, and to monitor the impact of implementation in terms of cost-effectiveness, access, and engagement, and the integration of models such as stepped care and clinical staging [11,30].

Fourth, it will be important to monitor the extent to which rapid responses to delivering mental health care in the context of the COVID-19 pandemic to address the increased need for mental health support among the general population [62] and those with existing MHA conditions [63-66] are reflected in mental health policy. Several of the P/Ts have responded to COVID-19 by providing information to the public about COVID-19 and maintaining mental health via social media [67]; delivering cognitive behaviour therapy through text messaging with mental health professionals (eg, Alberta's Text4Hope); expanding existing online mental health counseling programs (eg, British Columbia's MindHealthBC); and expanding use of existing mental health services for children and youth, such as text, live chat, mobile apps, and phone-based services [68]. At the same time, it is unclear whether the rapidly deployed initiatives are sufficient for responding to the mental health impacts of COVID-19 and to what extent will these initiatives address the already existing gaps in the continuum of mental health care for those with chronic mental health conditions. Given this, there is a continued need for comprehensive policy and research to support a coordinated and evidence-based integration of

XSL•FO
RenderX

technology in the Canadian mental health care system, during and after the COVID-19 pandemic.

Fifth, there is a further need for P/Ts and national collaboration to develop reliable and comprehensive environmental scans of the implementation of technology in Canadian mental health services, including analysis of failed attempts. Sixth, the development of a pan-Canadian framework for action that addresses the integration and coordination of technology in mental health service delivery can also help to guide P/T initiatives in this field.

Sixth, our findings provide a prepandemic baseline and replicable methods to inform a subsequent study of how the use of technology-supported services and innovations emerge as an important priority in mental health policy and practice, during and after the COVID-19 pandemic. Future research can also focus on examining how technology is considered in relation to the various mental health priorities addressed by policy documents (eg, use of technology in promoting wellness/recovery/MHA promotion, prevention, and early intervention; use of technology in promoting collaborative care; and use of technology in senior's mental health).

Limitations

We used an existing database (ie, the 19 P/T policy documents included in Sirotich et al [2]) as the main source of policy documents for our review. Therefore, it is possible that we might have missed some relevant documents. It is also acknowledged that there may be other documents potentially relevant to this review in these jurisdictions, but the documents we have identified outline government priorities for MHA care at the highest level and therefore are of primary importance for communicating a vision and strategy for mental health care services within a government's jurisdiction. It is also acknowledged that policies are frequently replaced; this review provides a report of relevant content available as of June 2019. In addition, we did not read through each page of each of the 22 documents to identify descriptions on the use of technology; although it was considered, it was deemed unfeasible by our team and also a potentially less reliable method of validation. We also believe that the final list of words we developed is quite comprehensive in its ability to capture the use of technology in mental health service delivery in these policy documents and that our methodology is both feasible and can be reliably replicated in future updated reviews.

Conclusions

This review on Canadian P/T mental health policy documents suggests that despite the potential for technology to improve access to and quality of MHA services, its implementation has been fragmented to date. Specifically, we identified large variations across Canadian P/T policy documents in terms of the types of technologies mentioned. Furthermore, many of the P/T policy documents mentioned only low-tech solutions, such as telephone-based interventions. The review also shows the limited attention given in the mental health policy documents to the evidence-based literature on the use of technology to deliver mental health services, suggesting a potential gap in knowledge on the part of policy makers regarding the evidence in this field. This gap can be reduced through collaborative interactions among academics, policy makers, and practitioners facilitated through sectors of P/T governments (eg, research and innovation, health services). It would be worth considering repeating this review at regular intervals, which would document the advancement of technology in care delivery, including related responses to the COVID-19 pandemic.

Acknowledgments

SL is funded by the Canada Research Chairs Program.

Authors' Contributions

Building on previous work conducted by CEA, SL conceived the plan and protocol for the study. Under the supervision of SL, LS extracted and analyzed data, and HL validated the data extraction and analyses. SL wrote the first draft of the manuscript. All authors contributed to revisions of this manuscript and approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of mental health policy documents reviewed.
[DOCX File , 31 KB - jmir_v23i5e25547_app1.docx]

Multimedia Appendix 2
Recurrent themes identified in mental health and addictions (MHA) policy priorities across the 13 Canadian provinces and territories.
[DOCX File , 22 KB - jmir_v23i5e25547_app2.docx]

Multimedia Appendix 3
Frequency of technology-related categories mentioned by each province or territory.

XSL•FO
RenderX

[DOCX File , 16 KB - jmir_v23i5e25547_app3.docx]

References

1. McDiarmid C. Health fact sheets: Mental health care needs, 2018. Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/n1/en/pub/82-625-x/2019001/article/00011-eng.pdf?st=gzDwQVNH [accessed 2021-04-23]
2. Sirotich F, Adair CE, Durbin J, Lin E, Canning C. Mental health and addictions policy priorities and performance measurement frameworks. Healthc Manage Forum 2019 Mar;32(2):105-112. [doi: 10.1177/0840470418810273] [Medline: 30739488]
3. Kozloff N, Mulsant BH, Stergiopoulos V, Voineskos AN. The COVID-19 Global Pandemic: Implications for People With Schizophrenia and Related Disorders. Schizophr Bull 2020 Jul 08;46(4):752-757 [FREE Full text] [doi: 10.1093/schbul/sbaa051] [Medline: 32343342]
4. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):1-1 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
5. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: Practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
6. Bierbooms JJPA, van Haaren M, IJsselsteijn WA, de Kort YAW, Feijt M, Bongers IMB. Integration of online treatment into the "new normal" in mental health care in post-COVID-19 times: Exploratory qualitative study. JMIR Form Res 2020 Oct 08;4(10):e21344 [FREE Full text] [doi: 10.2196/21344] [Medline: 33001835]
7. Chen JA, Chung W, Young SK, Tuttle MC, Collins MB, Darghouth SL, et al. COVID-19 and telepsychiatry: early outpatient experiences and implications for the future. Gen Hosp Psychiatry 2020;66:89-95 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.07.002] [Medline: 32750604]
8. 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. The Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
9. Christensen H, Griffiths K, Evans K. E-mental health in Australia: Implications of the Internet and related technologies for policy. Canberra: Commonwealth Department of Health and Ageing; 2002.
10. E-mental health in Canada: transforming the mental health system using technology. Mental Health Commission of Canada. URL: https://www.mentalhealthcommission.ca/sites/default/files/MHCC_E-Mental_Health-Briefing_Document_ENG_0.pdf [accessed 2021-04-23]
11. Lal S. E-mental health: Promising advancements in policy, research, and practice. Healthc Manage Forum 2019 Mar;32(2):56-62 [FREE Full text] [doi: 10.1177/0840470418818583] [Medline: 30739487]
12. Lal S, Adair CE. E-mental health: A rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
13. Krausz R, Ramsey D, Wetterlin F, Tabiova K, Thapliyal A. Accessible and cost-effective mental health care using e-mental health (EMH). In: Javed A, Fountoulakis K, editors. Advances in Psychiatry. Cham, Switzerland: Springer; 2019:129-141.
14. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: Evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
15. Liddy C, Rowan MS, Afkham A, Maranger J, Keely E. Building access to specialist care through e-consultation. Open Med 2013;7(1):e1-e8 [FREE Full text] [Medline: 23687533]
16. E-mental health: What's all the fuss about? NHS Confederation. URL: https://www.nhsconfed.org/resources/2013/01/e-mental-health-whats-all-the-fuss-about [accessed 2021-04-23]
17. Lester D. Crisis Intervention and Counseling by Telephone. Springfield, IL: Charles C Thomas Publisher; 2002.
18. 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]
19. Lal S, Abdel-Baki A, Sujanani S, Bourbeau F, Sahed I, Whitehead J. Perspectives of young adults on receiving telepsychiatry services in an urban early intervention program for first-episode psychosis: A cross-sectional, descriptive survey study. Front Psychiatry 2020;11:117 [FREE Full text] [doi: 10.3389/fpsyt.2020.00117] [Medline: 32194457]
20. 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]
21. Murray E. Web-based interventions for behavior change and self-management: Potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
22. Lal S, Gleeson J, Rivard L, D'Alfonso S, Joober R, Malla A, et al. Adaptation of a digital health innovation to prevent relapse and support recovery in youth receiving services for first-episode psychosis: Results from the Horyzons-Canada phase 1 study. JMIR Form Res 2020 Oct 29;4(10):e19887 [FREE Full text] [doi: 10.2196/19887] [Medline: 33118945]

XSL•FO
RenderX

23. 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]
24. De Crescenzo F, Economou A, Sharpley AL, Gormez A, Quested DJ. Actigraphic features of bipolar disorder: A systematic review and meta-analysis. Sleep Med Rev 2017 Jun;33:58-69. [doi: 10.1016/j.smrv.2016.05.003] [Medline: 28185811]
25. 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]
26. Wiederhold BK, Riva G. Virtual reality therapy: emerging topics and future challenges. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):3-6. [doi: 10.1089/cyber.2018.29136.bkw] [Medline: 30649958]
27. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
28. Linardon J, Shatte A, Messer M, Firth J, Fuller-Tyszkiewicz M. E-mental health interventions for the treatment and prevention of eating disorders: an updated systematic review and meta-analysis. J Consult Clin Psychol 2020 Nov;88(11):994-1007. [doi: 10.1037/ccp0000575] [Medline: 32852971]
29. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
30. McGrath P, Wozney L, Rathore S, Notarianni M, Schellenberg M. Toolkit for e-mental health implementation. Ottawa, ON: Mental Health Commission of Canada; 2018. URL: https://www.mentalhealthcommission.ca/sites/default/files/2018-09/E_Mental_Health_Implementation_Toolkit_2018_eng.pdf [accessed 2021-04-23]
31. Brownson RC, Jones E. Bridging the gap: Translating research into policy and practice. Prev Med 2009 Oct;49(4):313-315. [doi: 10.1016/j.ypmed.2009.06.008] [Medline: 19555708]
32. Gehring ND, McGrath P, Wozney L, Soleimani A, Bennett K, Hartling L, et al. Pediatric eMental healthcare technologies: a systematic review of implementation foci in research studies, and government and organizational documents. Implement Sci 2017 Jun 21;12(1):76 [FREE Full text] [doi: 10.1186/s13012-017-0608-6] [Medline: 28637479]
33. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: A systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
34. E-mental health strategy for Australia. Australian Government, Department of Health and Aging. 2012 Jun. URL: https://www1.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-e-emstrat-toc [accessed 2021-04-23]
35. Measuring progress: Resources for developing a mental health and addictions performance measurement framework for Canada. Mental Health Commission of Canada. 2018 Jul. URL: https://www.mentalhealthcommission.ca/sites/default/files/2018-07/MHPMF_final_report_eng_0.pdf [accessed 2021-04-23]
36. Sivumut abluqta: Stepping forward together. Government of Nunavut. 2014 Mar 20. URL: https://www.gov.nu.ca/sites/default/files/sivumut_abluqta_-_eng.pdf [accessed 2021-04-23]
37. A pathway to hope: a roadmap for making mental health and addictions care better for people in British Columbia. Government of British Columbia. 2019 Jun 26. URL: https://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/initiatives-plans-strategies/mental-health-and-addictions-strategy/bcmentalhealthroadmap_2019web-5.pdf [accessed 2021-04-23]
38. Together we can: progress update. Nova Scotia Department of Health and Wellness. 2016 Jan. URL: https://novascotia.ca/dhw/mental-health/reports/ProgressReport-Jan2016.pdf [accessed 2021-04-23]
39. Resiliency within an action plan for suicide prevention in Nunavut 2016-2017. Government of Nunavut. 2016 Mar. URL: https://www.gov.nu.ca/sites/default/files/resiliency_within_eng.pdf [accessed 2021-04-23]
40. Mental Health and Addictions Leadership Advisory Council. 2017. URL: http://www.health.gov.on.ca/en/common/ministry/publications/reports/bmhmbh_2017/vision_2017.pdf [accessed 2021-04-23]
41. B.C.'s mental health and substance use strategy 2017-2020. Government of British Columbia. 2017. URL: http://www.llbc.leg.bc.ca/public/pubdocs/bcdocs2017_2/681562/mental-health-substance-use-strategy.pdf [accessed 2021-04-23]
42. A path forward: BC first nations and aboriginal people's mental wellness and substance use 10-year plan. First Nations Health Authority, British Columbia Ministry of Health & Health Canada. 2013 Mar 18. URL: https://www.fnha.ca/Documents/FNHA_MWSU.pdf [accessed 2021-04-23]
43. Mulvale G, Kutcher S, Winkup J. A child and youth mental health and addiction framework for the Yukon. Government of Yukon. Hamilton, ON: McMaster University; 2014. URL: https://yawc.ca/files/a-child-youth-mental-health-addictions-framework-for-yukon-2014.pdf [accessed 2021-04-23]
44. 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]
45. Rising to the challenge: A strategic plan for the mental health and well-being of Manitobans. Government of Manitoba. 2011 Jun. URL: https://www.gov.mb.ca/health/mh/docs/challenge.pdf [accessed 2021-04-23]

XSL•FO
RenderX

46. New Brunswick family plan: Supporting those with addictions and mental health challenges. Government of New Brunswick. 2017 May. URL: http://www2.gnb.ca/content/dam/gnb/Departments/eco-bce/Promo/family_plan/mentalhealth.pdf [accessed 2021-04-23]
47. Mind and spirit: Promoting mental health and addictions recovery in the Northwest Territories, strategic framework 2016-2021. Government of Northwest Territories. 2016 Nov 03. URL: https://www.hss.gov.nt.ca/sites/hss/files/resources/mind-spirit-promoting-mental-health-addictions-recovery-nwt-strategic-framework-web.pdf [accessed 2021-04-23]
48. Moving forward together: Mental health and addiction strategy. Government of Prince Edward Island. 2016 Nov 14. URL: https://www.princeedwardisland.ca/sites/default/files/publications/peimentalhealthaddictionsstrategy_moving_forward.pdf [accessed 2021-04-23]
49. Faire ensemble et autrement: Plan d'action en santé mentale 2015-2020. Ministère de la Santé et des Services Sociaux. 2015 Oct 9. URL: https://publications.msss.gouv.qc.ca/msss/fichiers/2017/17-914-17W.pdf [accessed 2021-04-23]
50. Working together for change: A 10-year mental health and addictions action plan for Saskatchewan. Government of Saskatchewan. 2014 Dec 1. URL: https://www.saskatchewan.ca/government/health-care-administration-and-provider-resources/saskatchewan-health-initiatives/mental-health-and-addictions-action-plan [accessed 2021-04-23]
51. Valuing mental health: next steps. Alberta Government. 2017. URL: https://tinyurl.com/8awmmkkh [accessed 2021-04-23]
52. Valuing mental health: report of the Alberta mental health review committee 2015. Alberta Government. 2016 Feb 23. URL: https://open.alberta.ca/dataset/d8413604-15d1-4f15-a979-54a97db754d4/resource/1a5e7a16-3437-428e-b51f-4ba9201767a4/download/alberta-mental-health-review-2015.pdf [accessed 2021-04-23]
53. Towards recovery: A vision for a renewed mental health and addictions system for Newfoundland and Labrador. All-Party Committee on Mental Health and Addictions. 2017 Mar. URL: https://www.gov.nl.ca/hcs/files/all-party-committe-report.pdf [accessed 2021-04-23]
54. The way forward. Towards recovery: the mental health and addictions action plan for Newfoundland and Labrador. Government of Newfoundland and Labrador. 2017 Jun 30. URL: https://www.gov.nl.ca/hcs/files/mentalhealth-committee-mentalhealth-pdf-mentalhealth-addictions-plan.pdf [accessed 2021-04-23]
55. Together we can: the plan to improve mental health and addictions care for Nova Scotians. Nova Scotia Department of Health and Wellness. 2012 Apr 23. URL: https://novascotia.ca/dhw/mental-health/reports/Mental-Health-and-Addictions-Strategy-Together-We-Can.pdf [accessed 2021-04-23]
56. Mental Health and Addictions Leadership Advisory Council. 2015. URL: http://www.health.gov.on.ca/en/common/ministry/publications/reports/bmhmbh/mental_health_adv_council.pdf [accessed 2021-04-23]
57. Mental Health and Addictions Leadership Advisory Council. 2016. URL: http://www.health.gov.on.ca/en/common/ministry/publications/reports/bmhmbh_2016/moving_forward_2016.pdf [accessed 2021-04-23]
58. Open minds, healthy minds. Ontario's comprehensive mental health and addictions strategy. Government of Ontario. 2011 Jun. URL: http://www.health.gov.on.ca/en/common/ministry/publications/reports/mental_health2011/mentalhealth_rep2011.pdf [accessed 2021-04-23]
59. Forward together. Yukon mental wellness strategy 2016-2026. Yukon Health and Social Services. 2016 Jun. URL: https://yawc.ca/files/forward-together-yukon-mental-wellness-strategy-2016-26.pdf [accessed 2021-04-23]
60. Rush KL, Howlett L, Munro A, Burton L. Videoconference compared to telephone in healthcare delivery: a systematic review. Int J Med Inform 2018 Oct;118:44-53. [doi: 10.1016/j.ijmedinf.2018.07.007] [Medline: 30153920]
61. Choi BCK, Pang T, Lin V, Puska P, Sherman G, Goddard M, et al. Can scientists and policy makers work together? J Epidemiol Community Health 2005 Aug;59(8):632-637 [FREE Full text] [doi: 10.1136/jech.2004.031765] [Medline: 16020638]
62. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
63. Brown E, Gray R, Lo Monaco S, O'Donoghue B, Nelson B, Thompson A, et al. The potential impact of COVID-19 on psychosis: a rapid review of contemporary epidemic and pandemic research. Schizophr Res 2020 Aug;222:79-87 [FREE Full text] [doi: 10.1016/j.schres.2020.05.005] [Medline: 32389615]
64. Hawke LD, Barbic SP, Voineskos A, Szatmari P, Cleverley K, Hayes E, et al. Impacts of COVID-19 on youth mental health, substance use, and well-being: a rapid survey of clinical and community samples. Répercussions de la COVID-19 sur la santé mentale, l'utilisation de substances et le bien-être des adolescents: un sondage rapide d'échantillons cliniques et communautaires. Can J Psychiatry 2020 Oct;65(10):701-709 [FREE Full text] [doi: 10.1177/0706743720940562] [Medline: 32662303]
65. Anmella G, Arbelo N, Fico G, Murru A, Llach CD, Madero S, et al. COVID-19 inpatients with psychiatric disorders: real-world clinical recommendations from an expert team in consultation-liaison psychiatry. J Affect Disord 2020 Sep 01;274:1062-1067 [FREE Full text] [doi: 10.1016/j.jad.2020.05.149] [Medline: 32663933]
66. Warfield SC, Pack RP, Degenhardt L, Larney S, Bharat C, Ashrafioun L, et al. The next wave? Mental health comorbidities and patients with substance use disorders in under-resourced and rural areas. J Subst Abuse Treat 2021 Feb;121:108189. [doi: 10.1016/j.jsat.2020.108189] [Medline: 33162261]

XSL•FO
RenderX

67. Provincial and territorial resources for COVID-19. Government of Canada. 2020 Oct 16. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/symptoms/provincial-territorial-resources-covid-19.html [accessed 2021-04-23]
68. Provincial and territorial COVID-19 resources. Mental Health Commission of Canada. 2020. URL: https://www.mentalhealthcommission.ca/English/provincial-and-territorial-covid-19-resources [accessed 2021-04-23]

Abbreviations

AAOR: Service d'accueil, d'analyse, d'orientation et de référence des services sociaux généraux
CNESM: Centre national d'excellence en santé mentale
EHR: electronic health record
EMR: electronic medical record
ICT: information and communication technology
iCBT: internet cognitive behavioural therapy
MHA: mental health and addictions
MHCC: Mental Health Commission of Canada's
P/T: provincial/territorial

Edited by G Eysenbach; submitted 10.11.20; peer-reviewed by J Bierbooms, H LaMonica, Y Wang, M Newton, X Lv, M Dyson; comments to author 02.12.20; revised version received 01.03.21; accepted 04.04.21; published 05.05.21.

Please cite as:
Lal S, Siafa L, Lee H, Adair CE
Priority Given to Technology in Government-Based Mental Health and Addictions Vision and Strategy Documents: Systematic Policy Review
J Med Internet Res 2021;23(5):e25547
URL: https://www.jmir.org/2021/5/e25547
doi:10.2196/25547
PMID:33949955

©Shalini Lal, Lyna Siafa, Hajin Lee, Carol E Adair. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2021. 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 Patterns and Impact of Social Media Exposure of Journal Publications in Gastroenterology: Retrospective Cohort Study

Austin Lee Chiang[1*], MD, MPH; Loren Galler Rabinowitz[2*], MD; Javid Alakbarli[3*], MD; Walter W Chan[1*], MD, MPH

[1]Brigham and Women's Hospital, Boston, MA, United States

[2]Columbia University Medical Center, New York, NY, United States

[3]Beth Israel Deaconess Medical Center, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Austin Lee Chiang, MD, MPH
Brigham and Women's Hospital
75 Francis St
Boston, MA
United States
Phone: 1 617 732 4840
Email: austinleechiang@gmail.com

Abstract

Background: Medical journals increasingly promote published content through social media platforms such as Twitter. However, gastroenterology journals still rank below average in social media engagement.

Objective: We aimed to determine the engagement patterns of publications in gastroenterology journals on Twitter and evaluate the impact of tweets on citations.

Methods: This was a retrospective cohort study comparing the 3-year citations of all full-length articles published in five major gastroenterology journals from January 1, 2012, to December 31, 2012, tweeted by official journal accounts with those that were not. Multivariate analysis using linear regression was performed to control for journal impact factor, time since publication, article type, frequency of reposting by other users ("retweets"), and media addition to tweets. Secondary analyses were performed to assess the associations between article type or subtopic and the likelihood of social media promotion/engagement.

Results: A total of 1666 articles were reviewed, with 477 tweeted by the official journal account. Tweeting an article independently predicted increased citations after controlling for potential confounders (β coefficient=13.09; *P*=.007). There was significant association between article type and number of retweets on analysis of variance (ANOVA) (*P*<.001), with guidelines/technical reviews (mean difference 1.04, 95% CI 0.22-1.87; *P*<.001) and meta-analyses/systemic reviews (mean difference 1.03, 95% CI 0.35-1.70; *P*<.001) being retweeted more than basic science articles. The manuscript subtopics most frequently promoted included motility/functional bowel disease (odds ratio [OR] 3.84, 95% CI 1.93-7.64; *P*<.001) and education (OR 4.69, 95% CI 1.62-13.58; *P*=.004), while basic science papers were less likely tweeted (OR 0.154, 95% CI 0.07-0.34; *P*<.001).

Conclusions: Tweeting of gastroenterology journal articles independently predicted higher 3-year citations. Wider adoption of social media to increase reach and measure uptake of published research should be considered.

(J Med Internet Res 2021;23(5):e25252) doi:10.2196/25252

KEYWORDS

social media; gastroenterology journals; gastroenterology research; journal citations

Introduction

Social media is playing an increasingly important role in health care as an inexpensive way to improve accessibility of medical information. In recent years, medical journals have created accounts on social media to share published content and improve visibility among both mainstream audiences and health professionals. Unlike traditional media, social media differs in its ability to curate content (using tools like hashtags) and to facilitate engagement by readers and viewers. It is often challenging to stay abreast of all newly published data from the multitude of scientific journals within a given medical specialty.

XSL•FO
RenderX

Social media promotion may help narrow the range of what is considered most relevant to target audiences and/or newsworthy.

Twitter is the prime social media platform for online discussion with 335 million active monthly users worldwide and more than 500 million "tweets" per day [1]. The content of each tweet is restricted to 280 characters and may contain links to external websites. Each tweet is visible to followers of the account, and individuals can engage by reposting (ie, "retweeting") content in their own Twitter feeds or "liking" posts.

Twitter activity may predict publication and overall journal performance. A prior study found a significant association between Twitter followers of an official journal account and both the journal impact factor and total citations [2], with an estimated 1% increase in journal citations for every 0.62% increase in Twitter followers. Among adult and pediatric urology journals, having an official Twitter account was found to be associated with a greater impact factor [3,4]. Following a targeted effort to promote published articles to their 700 followers, the *Journal of Neurointerventional Surgery* saw a "substantial increase" of 1500 visits to its scientific content [5].

The effects of social media exposure of published research in gastroenterology remain unclear. More specifically, the impacts on citations and patterns of social media promotion of these publications are also unknown. While the process that journals employ to select and promote articles on social media is largely opaque and likely not random, we may be able to glean the patterns of social media amplification and control for confounders like manuscript type and subject matter. Characterizing these patterns may serve to highlight areas that are lacking exposure or reflect what journals perceive as most relevant to the general public.

The primary aim of this study was to determine if social media promotion by gastroenterology journal Twitter accounts is associated with greater number of citations. We also aimed to determine if there is preferential promotion of certain types of publications and subtopics by gastroenterology journals on Twitter.

Methods

This was a retrospective cohort study assessing the citations of all full-length articles published in five major peer-reviewed gastroenterology journals (*American Journal of Gastroenterology*, *Clinical Gastroenterology and Hepatology*, *Gastroenterology*, *Gastrointestinal Endoscopy*, and *Pancreas*) from January 1, 2012, to December 31, 2012. These journals and this date range were selected for the following reasons: to allow sufficient time for citations to accrue, to account for the diminishing impact of a tweet over time, and to select a year where all five journals appeared on social media.

The number of citations in academic literature as of November 15, 2015, according to Google Scholar was compared between articles tweeted by the official journal accounts and those that were not. Google Scholar has been previously used in similar research to quantify citations [6]. Publications were further categorized by manuscript type and subtopic. Manuscript types included prospective research, retrospective research, basic science, meta-analyses and systematic reviews, guidelines and technical reviews, case reports, video publications, and editorials. Manuscript subtopics included esophagus, gastric, small bowel, colon, liver, pancreas, biliary, motility and functional bowel disease, cancer, basic science, quality improvement, cost-effectiveness, inflammatory bowel disease, endoscopy, education, and microbiome. Multiple assignments for subtopics were permitted.

The mean number of citations between tweeted articles and those that were not tweeted by the official journal accounts were compared using Student *t* test. To detect an independent association between social media promotion on Twitter and Google Scholar citations, multivariate analysis using linear regression was performed to control for the 2012 journal impact factor (as published by Thomson Reuters), time since publication, and article type. Among articles that were tweeted, the overall associations between manuscript types and number of citations were assessed using analysis of variance (ANOVA). Pairwise comparisons of manuscript types with regard to number of citations were performed using the Bonferroni method.

To evaluate the likelihood of specific subtopics being promoted on social media by gastroenterology journals, the rates of tweeted manuscripts for each subtopic were compared to those not containing the corresponding subtopic using the chi-squared test. Multivariate analysis was performed using logistic regression, controlling for potential confounders including the specific journal, manuscript type, and number of citations. Results of regression analyses are expressed as raw coefficients to demonstrate the effect of each variable on citations and to better generalize these results to other settings or populations.

All statistical analyses were performed using SAS Version 9.4 (SAS Institute).

Results

In total, 1666 gastroenterology articles were reviewed, with 477 having been tweeted by official journal accounts. In 2012, *Gastrointestinal Endoscopy* published 451 articles, *Pancreas* published 242 articles, *American Journal of Gastroenterology* published 226 articles, *Gastroenterology* published 473 articles, and *Clinical Gastroenterology and Hepatology* published 274 articles (Table 1). On univariate analysis, articles that were tweeted had a significantly higher number of citations compared with nontweeted articles (36.9 vs 27.4, *P*=.04) (Figure 1). On multivariate analysis, tweeting of an article (β coefficient=13.09, *P*=.007) was independently associated with increased citations after controlling for potential confounders (Table 2). Not surprisingly, the duration since publication (in days) was found to be a predictor for increased citations in the linear regression model. Among tweeted articles, those that were retweeted (a possible proxy for strong public interest or colleague endorsement) also had higher citations compared with those that were not retweeted (72.3 vs 17.6, *P*=.004) (Figure 1), although this did not reach statistical significance in the multivariate model (β coefficient=23.2, *P*=.28), likely due to a small sample size.

XSL•FO
RenderX

Table 1. Characteristics of the articles in the five gastroenterology journals (Twitter promotion, impact factor, total citations, and frequency of publication by manuscript type and subtopic).

Characteristic	Journal					Total (N=1424)
	AJG[a] (N=226)	*CGH*[b] (N=274)	*Gastro*[c] (N=473)	*GIE*[d] (N=451)	*Pancreas* (N=242)	
Tweeted, n	31	69	70	137	170	477
Not tweeted, n	195	205	403	314	72	1189
2012 impact factor	7.282	5.627	11.675	4.878	2.386	N/A[e]
Total citations, n	11,286	6008	21,264	8336	3201	50,095
Manuscript type, n						
Prospective studies	77	38	50	118	59	342
Retrospective studies	54	86	35	102	42	319
Basic science studies	10	1	159	6	106	282
Meta-analyses, systematic reviews	39	26	37	23	14	139
Guidelines, technical reviews	7	18	4	25	2	56
Editorials	34	38	58	42	5	177
Case reports	1	67	129	122	14	333
Videos	4	0	1	13	0	18
Publication subtopic, n						
Pediatric	6	4	8	9	2	29
Esophagus	29	38	33	55	0	155
Gastric	14	12	33	44	1	104
Small bowel	22	24	27	35	0	108
Colon	53	47	65	94	0	259
Liver	26	75	151	10	2	264
Pancreas	10	28	40	59	234	371
Biliary	8	19	19	47	1	94
Motility/functional	35	9	8	5	0	57
Cancer	22	43	89	85	124	363
Basis science	11	3	189	3	109	315
Quality improvement	19	24	21	47	4	115
Cost-effectiveness	1	6	2	0	1	10
Inflammatory bowel disease	30	26	32	4	0	92
Endoscopy	35	59	27	362	14	497
Education	2	1	9	7	0	19
Infectious disease/microbiome	16	7	16	3	0	42

[a]AJG: American Journal of Gastroenterology.

[b]CGH: Clinical Gastroenterology and Hepatology.

[c]Gastro: Gastroenterology.

[d]GIE: Gastrointestinal Endoscopy.

[e]N/A: not applicable.

XSL•FO
RenderX

Figure 1. Comparison of citations between articles that were tweeted and those that were not tweeted. Primary tweets: analysis of all manuscripts (n=1666) comparing tweeted articles and nontweeted articles; retweets: analysis of all tweeted manuscripts (n=477) comparing articles that were retweeted at least once and articles without retweets.

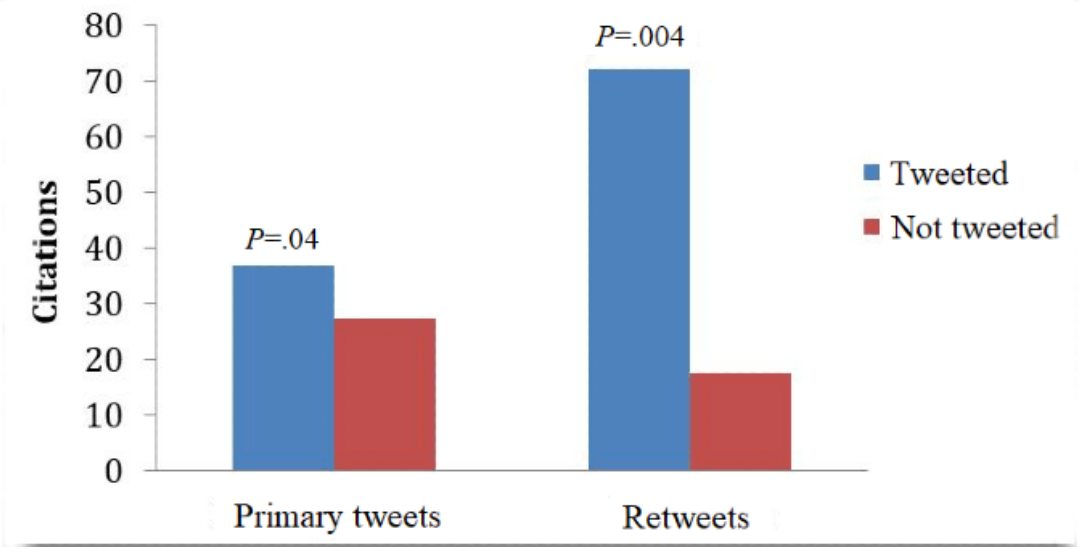


Table 2. Multivariate linear regression analysis assessing the predictors of the number of Google Scholar citations among all articles published by five major gastroenterology journals in 2012 (N=1666, tweeted n=477).

Variable	β coefficient (SE)	*P* value
Tweeted article	13.09 (4.82)	.007
Time since publication (days)	0.06 (0.02)	.003
Journal		
GIE[a]	Reference	Reference
Gastro[b]	37.8 (5.65)	<.001
AJG[c]	23.8 (6.71)	<.001
CGH[d]	3.18 (6.12)	.60
Pancreas	−12.2 (7.12)	.09
Manuscript type		
Prospective studies	Reference	Reference
Retrospective studies	−8.88 (6.20)	.15
Basic science studies	−23.5 (7.02)	<.001
Meta-analyses/systematic reviews	26.0 (7.98)	.001
Guidelines, technical reviews	52.8 (11.5)	<.001
Editorials	−41.9 (7.42)	<.001
Case reports	−43.2 (6.40)	<.001
Videos	−21.8 (19.2)	.26

[a]GIE: Gastrointestinal Endoscopy.

[b]Gastro: Gastroenterology.

[c]AJG: American Journal of Gastroenterology.

[d]CGH: Clinical Gastroenterology and Hepatology.

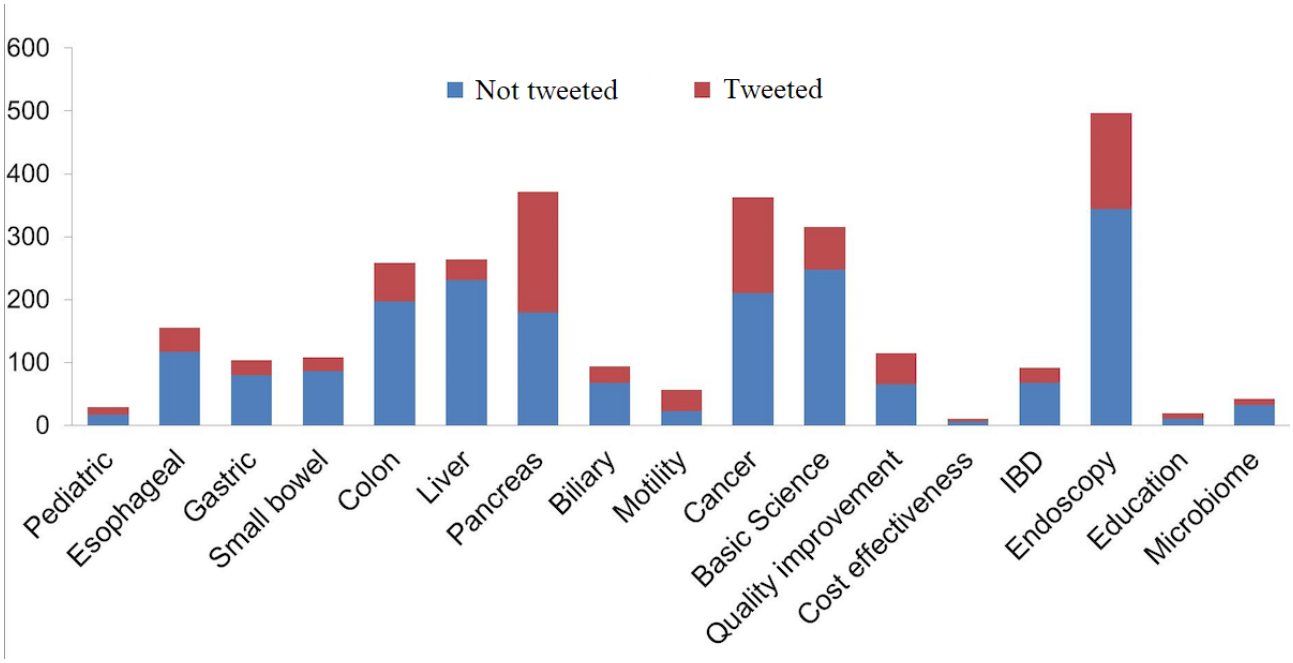
On univariate analysis, articles identified in the categories of pancreas (odds ratio [OR] 3.80, *P*<.001), cancer (OR 2.20, *P*<.001), and quality improvement (OR 2.05, *P*=.003) were associated with increased Twitter promotion, while small bowel (OR 0.62, *P*=.05) and basic science (OR 0.63, *P*=.002) articles were associated with less social media exposure.

After controlling for other covariates, including other manuscript subtypes, on multivariate analysis, the subtopics of motility and functional bowel disease (OR 3.84, 95% CI 1.93-7.64, *P*<.001), cancer (OR 1.392, 95% CI 1.016-1.909, *P*=.04), education (OR 4.69, 95% CI 1.62-13.58, *P*=.004), and quality improvement (OR 2.41, 95% CI 1.52-3.84, *P*<.001) were independently

XSL•FO
RenderX

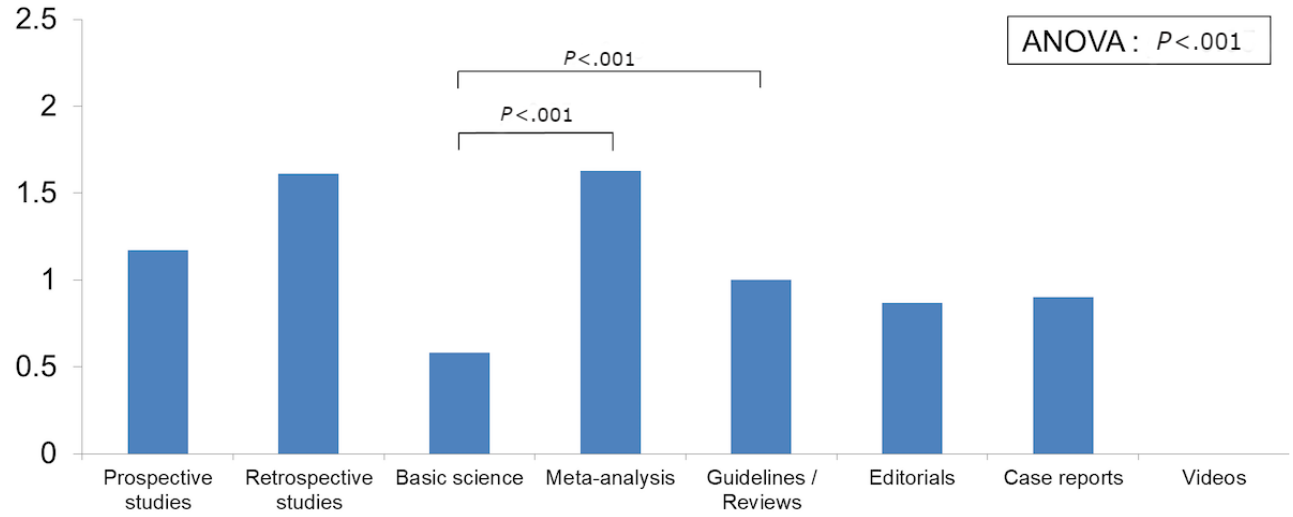
associated with increased promotion on Twitter, while basic science articles were significantly less likely to be tweeted (OR 0.154, 95% CI 0.07-0.34, *P*<.001) (Figure 2).

Figure 2. Number of publications (both tweeted and nontweeted) in 2012 by subtopic among the five journals included. IBD: inflammatory bowel disease.



A significant overall association between article type and citations was noted on ANOVA (*P*<.001), with guidelines/technical reviews (mean 90.23, SD 355.57), meta-analyses/systemic reviews (mean 75.22, SD 124.85), and prospective studies (mean 42.96, SD 54.26) having the most citations. On multivariate analysis using prospective studies as a reference, guidelines/technical reviews (β coefficient=52.8, *P*<.001) and meta-analyses/systemic reviews (β coefficient=26, *P*=.001) were significantly more likely to be cited, while basic science articles (β coefficient=−23.5, *P*<.001), case reports (β coefficient=−43.2, *P*<.001), and editorials (β coefficient=−41.9, *P*<.001) had significantly fewer citations (Table 2). Among tweeted articles, there was a significant association between article type and number of retweets on ANOVA (*P*<.001). On pairwise comparison, guidelines/technical reviews (mean difference 1.04, 95% CI 0.22-1.87, *P*<.001) and meta-analyses/systemic reviews (mean difference 1.04, 95% CI 0.22-1.86, *P*<.001) were being retweeted significantly more than basic science articles (Figure 3). Media additions to tweets were not associated with the number of citations or retweets.

Figure 3. Mean number of retweets by article type.



XSL•FO
RenderX

Discussion

Principal Findings

Our study showed that social media promotion of gastroenterology journal publications on Twitter independently predicted a greater number of Google Scholar citations in 3 years after controlling for journal impact factor, type of article, and time since publication.

We also found that tweeted articles that were retweeted had significantly increased citations, perhaps a reflection that public engagement with these tweets might predict further academic interest. Our study also demonstrated that guidelines/technical reviews and meta-analyses/systematic reviews are among the most heavily promoted article types by official journal accounts on Twitter. Articles involving motility/neurogastroenterology, cancer, education, and quality improvement were independently associated with increased social media exposure, while basic science manuscripts were less promoted. This may reflect the perceived areas of public interest in gastroenterology and influence perception of gastroenterology research, resulting in potential secondary impacts on factors, such as funding. Further studies are needed to evaluate why these promotional differences exist and how this might affect research funding availability, public interest, and clinical and epidemiological outcomes.

As done in studies from other disciplines, Twitter was chosen given that it is the predominant mainstream social media platform where journal publications are regularly promoted, where academic discussion exists, and where analytics are most readily available. The evidence regarding the association between Twitter promotion and publication citations has varied across different disciplines. One study of 20 ecological journals showed a similar positive correlation between social media exposure and citations, independent of time since publication and impact factor [7]. In the *American Journal of Psychiatry*, a study of 438 published articles suggested that a greater frequency of Twitter mentions was associated with more citations [8]. One study of 286 articles from the *Journal of Medical Internet Research* showed that highly tweeted articles were 11 times more likely to be highly cited and that Twitter mentions in the first 3 days could predict highly cited articles [6].

However, this observation was not consistent across all disciplines. In a randomized trial of social media impact on 243 publications from *Circulation*, there was no significant association between social media exposure and 30-day article website views according to Google analytic data [9]. There were also no differences noted by article subtype (referring to the general categories of clinical, population, or basic science articles). One study of all 1.4 million biomedical publications between 2010 and 2012 found a weak correlation between Twitter mentions of journal articles and traditional bibliometric indicators, although the authors concluded that Twitter mentions did not reflect traditional research impact [10].

The number of annual citations by a journal is used to calculate a journal's impact factor [11]. Historically, journal impact factor has been directly correlated with citations, as impact factor is the ratio of the number of citations to the number of publications by a journal in a given year. Google Scholar was chosen as the source of citations given its inclusivity of citations among all sources, unlike Web of Science and Scopus, which only consider citations from within journals listed in their libraries. However, some researchers have raised concerns that certain tactics can be used to boost the journal impact factor, such as publishing more review articles, strategic publication timing, and internal citations. Citations also take time to accumulate. Social media may therefore play another role in measuring the societal impact of an individual article. As a result, other metrics collectively known as "altmetrics," which include not only number of citations, but also social media views and engagement, references by databases, and news media, are emerging [12].

In our study, guidelines and reviews published in major gastroenterology journals were the most heavily promoted types of manuscripts on social media. This promotional pattern may be a result of known interest in these articles, given existing knowledge that guidelines and reviews accrue more citations than other publication types, and they may be more relevant to a broader audience that includes general clinicians and practitioners [13]. For gastrointestinal subtopics, our finding of functional bowel disease and cancer articles as the most tweeted is also not surprising. Communities for these conditions are very active on social media, and journals may be preferentially promoting these articles knowing the discussion and engagement they tend to generate. Moreover, the high general prevalence of functional bowel conditions and the increased interest cancer-related topics tend to generate in the lay media may also play roles in the higher social media activities of these articles.

Social media engagement by the gastroenterology community is lagging behind other medical subspecialties. Adoption of Twitter by journals still varies widely by specialty, with 70% of the 20 leading radiology journals on Twitter, but only 28.1% of general medical journals and 39% of urological journals on Twitter [2,14]. Latest data suggest that the proportion of overall tweeted gastroenterology publications (7.8%) falls below the average rate of Twitter promotion (9.4%) of all medical journal articles [10]. A survey of 265 gastroenterologists in 2015 found that 82.1% of respondents did not access social media for journals or other educational purposes, and 47.7% reported never having used any form of social media. This is in contrast to patient acceptance of social media, where 84.4% (n=112) of inflammatory bowel disease patients and 72.9% (n=68) of chronic viral hepatitis patients favored interaction with health care professionals on social media [15]. This dynamic may be improving in recent years, as there has been a growing presence of gastroenterologists and young physicians on social media.

One limitation of this study is the inability to establish causality and elucidate the potential mechanisms of social media impact on citations. Articles addressing subjects of higher interest or "more popular" topics (therefore, more highly cited) may be selectively tweeted at a higher frequency. For instance, social media promotion has been associated with greater downloads of journal articles in clinical pain sciences, which could suggest greater readership as a catalyst for future citations [16]. The radiology community has observed a greater distribution of web links to published articles on Twitter [14]. In fact, the effect of

XSL•FO
RenderX

social media on journal article readership has been previously demonstrated in a study of the *Journal of the American College of Radiology*, which showed that a planned Twitter-based discussion increased monthly website journal article views by 31.4%, unique visitors by 20.0%, and website visits by 25.5% [17]. In a randomized controlled study of the same journal, Twitter promotion of publications was independently associated with significantly greater weekly webpage visits (18.2 vs 7.6 page visits) [17]. Another study showed that 7 out of 11 radiology journals new to Twitter experienced increases in impact factor after 1 year [14].

Moreover, it is unknown if social media alone can be credited for greater citations or if social media is simply one arm of a larger promotional effort that includes other avenues, such as press releases for traditional media coverage. It is worth noting, however, that because the initial promotional tweet precedes article downloads and subsequent citations, it is not possible for the number of citations to cause a reactionary increase in social media presence for that particular article. Additionally, various journals may adopt different methods to selectively promote certain subtopics or manuscript types (such as guidelines) on social media. To best account for these biases, we controlled for subtopics and manuscript types in the multivariate analysis. Another potential limitation was the potential overestimation of baseline academic impact using Google Scholar citations. Though Google Scholar may include duplicate citations or citations of a paper in a nonpeer-reviewed publication without scholarly relevance, the Google algorithm is standard and therefore objectively compares citation volume.

Conclusion

In conclusion, social media promotion on Twitter of gastroenterology publications independently predicted a greater number of citations in 3 years in this exploratory study. Publications and researchers should consider wider adoption of social media to increase reach and measure uptake of published research. Social media promotion of publications can not only potentially boost journal citations, but also help define the societal impact of an individual article and thus influence academic promotion. However, beyond citations and academic uptake, journals and physicians should be aware of other benefits of social media for professionals, patients, and family members. In addition to academic productivity, it is important to recognize how social media could have other important roles in public health. For instance, social media may help propel academic medicine by informing other professionals. Moreover, it is our public health responsibility as a medical community to serve as primary sources of accurate up-to-date medical information online in order to preserve the integrity of what readers consume. Internally, social media can provide an open forum for discussion, boost professional and institutional recognition, attract referrals for trial enrollment and research purposes, and perhaps encourage funding to sustain academic research.

Authors' Contributions

ALC: study concept/design, data collection, and drafting of the manuscript; LGR: data collection and drafting of the manuscript; JA: data collection; WC: study concept/design, data collection, statistical analysis, drafting/revision of the manuscript, and study supervision.

Conflicts of Interest

None declared.

References

1. Q2 2018 Selected Company Financials and Metrics. Twitter. URL: https://s22.q4cdn.com/826641620/files/doc_financials/2018/q2/Q2_2018_Selected_Company_Metrics_and_Financials.pdf [accessed 2021-04-08]
2. Cosco TD. Medical journals, impact and social media: an ecological study of the Twittersphere. CMAJ 2015 Dec 08;187(18):1353-1357 [FREE Full text] [doi: 10.1503/cmaj.150976] [Medline: 26644544]
3. 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):513.e1-513.e7. [doi: 10.1016/j.jpurol.2017.03.027] [Medline: 28483467]
4. 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]
5. Fargen KM, Ducruet AF, Hyer M, Hirsch JA, Tarr RW. Expanding the social media presence of the Journal of Neurointerventional Surgery: editor's report. J Neurointerv Surg 2017 Feb 29;9(2):215-218. [doi: 10.1136/neurintsurg-2015-012251] [Medline: 26927814]
6. 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]
7. Peoples BK, Midway SR, Sackett D, Lynch A, Cooney PB. Twitter Predicts Citation Rates of Ecological Research. PLoS One 2016;11(11):e0166570 [FREE Full text] [doi: 10.1371/journal.pone.0166570] [Medline: 27835703]
8. Quintana DS, Doan NT. Twitter Article Mentions and Citations: An Exploratory Analysis of Publications in the American Journal of Psychiatry. Am J Psychiatry 2016 Feb 01;173(2):194. [doi: 10.1176/appi.ajp.2015.15101341] [Medline: 26844798]

XSL•FO
RenderX

9. Fox CS, 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 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
10. Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. J Assn Inf Sci Tec 2013 Nov 26;65(4):656-669. [doi: 10.1002/asi.23101]
11. Garfield E. The history and meaning of the journal impact factor. JAMA 2006 Jan 04;295(1):90-93. [doi: 10.1001/jama.295.1.90] [Medline: 16391221]
12. Scarlat MM, Mavrogenis AF, Pećina M, Niculescu M. Impact and alternative metrics for medical publishing: our experience with International Orthopaedics. International Orthopaedics (SICOT) 2015 May 7;39(8):1459-1464. [doi: 10.1007/s00264-015-2766-y]
13. Haustein S, Costas R, Larivière V. Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. PLoS One 2015 Mar 17;10(3):e0120495 [FREE Full text] [doi: 10.1371/journal.pone.0120495] [Medline: 25780916]
14. Kelly BS, 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]
15. Timms C, Forton D, Poullis A. Social media use in patients with inflammatory bowel disease and chronic viral hepatitis. Clin Med (Lond) 2014 Apr 08;14(2):215-215 [FREE Full text] [doi: 10.7861/clinmedicine.14-2-215] [Medline: 24715144]
16. 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;8(7):e68914 [FREE Full text] [doi: 10.1371/journal.pone.0068914] [Medline: 23874810]
17. 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]

Abbreviations

ANOVA: analysis of variance
OR: odds ratio

Edited by R Kukafka; submitted 25.10.20; peer-reviewed by D Bowen, R Hawkins; comments to author 07.12.20; revised version received 09.12.20; accepted 11.03.21; published 14.05.21.

Please cite as:
Chiang AL, Rabinowitz LG, Alakbarli J, Chan WW
The Patterns and Impact of Social Media Exposure of Journal Publications in Gastroenterology: Retrospective Cohort Study
J Med Internet Res 2021;23(5):e25252
URL: https://www.jmir.org/2021/5/e25252
doi:10.2196/25252
PMID:33707166

©Austin Lee Chiang, Loren Galler Rabinowitz, Javid Alakbarli, Walter W Chan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.05.2021. 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

Evaluating Scholars' Impact and Influence: Cross-sectional Study of the Correlation Between a Novel Social Media–Based Score and an Author-Level Citation Metric

Lucas Oliveira J e Silva[1], MD; Graciela Maldonado[1], MD, MSc; Tara Brigham[2], MLS; Aidan F Mullan[3], MA; Audun Utengen[4], BSc, MBA; Daniel Cabrera[1], MD

[1]Department of Emergency Medicine, Mayo Clinic Rochester, Rochester, MN, United States

[2]Mayo Clinic Libraries, Mayo Clinic Florida, Jacksonville, FL, United States

[3]Department of Biostatistics and Informatics, Mayo Clinic Rochester, Rochester, MN, United States

[4]Symplur, Los Angeles, CA, United States

Corresponding Author:
Daniel Cabrera, MD
Department of Emergency Medicine
Mayo Clinic Rochester
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 255 4399
Email: cabrera.daniel@mayo.edu

Abstract

Background: The development of an author-level complementary metric could play a role in the process of academic promotion through objective evaluation of scholars' influence and impact.

Objective: The objective of this study was to evaluate the correlation between the Healthcare Social Graph (HSG) score, a novel social media influence and impact metric, and the h-index, a traditional author-level metric.

Methods: This was a cross-sectional study of health care stakeholders with a social media presence randomly sampled from the Symplur database in May 2020. We performed stratified random sampling to obtain a representative sample with all strata of HSG scores. We manually queried the h-index in two reference-based databases (Scopus and Google Scholar). Continuous features (HSG score and h-index) from the included profiles were summarized as the median and IQR. We calculated the Spearman correlation coefficients (ρ) to evaluate the correlation between the HSG scores and h-indexes obtained from Google Scholar and Scopus.

Results: A total of 286 (31.2%) of the 917 stakeholders had a Google Scholar h-index available. The median HSG score for these profiles was 61.1 (IQR 48.2), and the median h-index was 14.5 (IQR 26.0). For the 286 subjects with the HSG score and Google Scholar h-index available, the Spearman correlation coefficient ρ was 0.1979 (P<.001), indicating a weak positive correlation between these two metrics. A total of 715 (78%) of 917 stakeholders had a Scopus h-index available. The median HSG score for these profiles was 57.6 (IQR 46.4), and the median h-index was 7 (IQR 16). For the 715 subjects with the HSG score and Scopus h-index available, ρ was 0.2173 (P<.001), also indicating a weak positive correlation.

Conclusions: We found a weak positive correlation between a novel author-level complementary metric and the h-index. More than a chiasm between traditional citation metrics and novel social media–based metrics, our findings point toward a bridge between the two domains.

(J Med Internet Res 2021;23(5):e28859) doi:10.2196/28859

KEYWORDS

social media; Twitter; journal impact factor; h-index; digital scholarship; digital platform; Scopus; metrics; scientometrics; altmetrics; metrics; stakeholders; health care; digital health care

XSL•FO
RenderX

Introduction

Since the development of social media platforms and new communication channels, the use of traditional bibliographic metrics (ie, citation counts, h-indexes) as the predominant factors for academic performance has been questioned [1]. Traditional benchmarks such as citation counts fail to capture the authors' impact outside academic circles [2]. The ways in which research output is indexed, searched, located, read, and mentioned have significantly changed, and these ways do not describe the influence and impact that scholarly work may have outside core academic domains [3,4].

In the health care world, social media platforms (eg, Twitter, Facebook) are consistently used by patients, policy makers, clinicians, and researchers as efficient ways of sharing information, staying up to date with scientific knowledge, and collaborating with peers and patients [5]. The widespread use of social media by health care stakeholders has led to the development of alternative impact metrics, also known as "altmetrics" [6]. The altmetrics approach offers new ways to analyze and inform scholarship [7]. It complements rather than replaces traditional indicators of a scholar's performance [8]. Altmetrics have even been adopted aggressively by traditional publishing companies [9]. The study of these alternative metrics is an emerging field; unlike traditional parameters, such as the impact factor or h-index, it does not rely solely on citation counts but is a composite measure. It considers other features such as the number of knowledge databases that refer to the work, and the number of times the work has been viewed and downloaded; it also factors in the number of mentions in social media and traditional news outputs.

Academic merit and achievement should be appraised using frameworks such as the comprehensive researcher achievement model (CRAM) [8], encompassing a combination of four aspects: quantity of researcher outputs (productivity), value of outputs (quality), outcomes of research outputs (impact), and relations between publications or authors and the wider world (influence). Current traditional benchmarks focus mostly on productivity and quality, while alternative metrics focus on influence and impact. In 2011, Eysenbach proposed the Twimpact Factor, an article level social media impact metric consisting of the absolute cumulative number of tweetations 7 days after publication of the article, and the Twindex, which is the relative percentile of the Twimpact Factor of a given article compared with other articles in the same journal [10]. For articles published in the *Journal of Medical Internet Research*, Eysenbach found relatively strong article-level correlations between these metrics (number of tweets, adjusted by time and journal factors) and future citations and highlighted the importance of using social media–based impact measures to complement traditional citation metrics [10]. While social media metrics at the article or journal level already exist and have been correlated with traditional citation metrics [10], novel tools could also be used to evaluate features such as influence and impact at the author level. There is a clear need to improve the ways in which the different outputs of scholarly work are evaluated, as claimed by the Declaration on Research Assessment (DORA) movement [11]. The development of an author-level complementary metric could play a role in the academic promotion process through objective evaluations of scholars' influence and impact.

Recently, multiple organizations have created tools that attempt to measure influence and impact in the digital domain primarily by using network analysis of social media activity and digital publications [12]. Among these innovations, Symplur's Healthcare Social Graph (HSG) score has recently emerged [13]. In this context, we aimed to evaluate the correlation between the HSG, a social media influence and impact metric, and the h-index, a traditional author-level metric.

Methods

Study Design, Study Setting, and Participants

This report was written following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [14]. This study was deemed exempt by the Institutional Review Board.

This was a cross-sectional observational study of health care stakeholders with a social media presence randomly sampled from the HSG database in May 2020. Health care stakeholders included the following three taxonomic categories: "doctor" (ie, those identified as possibly licensed, MDs, DOs, PhDs), "health care professionals" (ie, those identified as other health care professionals such as nurses, dietitians, respiratory therapists, and pharmacists), and "researchers/academicians" (ie, people working in the field of health-related research or academia). Over 1 million Twitter profiles were labeled according to the health care stakeholder category as part of the database metadata. Only the profiles of those identifying themselves in their public Twitter profile received a label by Symplur partly through manual verification and partly through a machine learning process [15]. We did not exclude health care stakeholders based on their discipline. Considering the 6 million Twitter accounts that received an HSG score and individuals identified as health care stakeholders in May 2020, we performed stratified random sampling to obtain a representative sample with all strata of HSG scores. A random sample of 100 profiles from each HSG score decile (0-9, 10-19, etc) was obtained, yielding an initial list of 1000 subjects with their respective HSG scores. This stratification method was chosen owing to the skewness of the HSG scores in the Symplur database, where simple probability sampling would lead to a study population restricted to lower values of the HSG score.

Data Source, Variables, and Measurement

Symplur is a health care social media analytics company that created the HSG database holding public digital content (ie, conversations, interactions) originating from Twitter and obtained via the official Twitter application programming interface (API) while supplementing it with other public content from social media platforms including LinkedIn, YouTube, Instagram, Reddit, and Facebook. The HSG score was developed by Symplur to identify and rank influencers in any health care topic and is conceptually like an eigenvector [16]. This score ranks Twitter accounts by their global conversational impact in healthcare over the last 52 weeks. As long as Twitter accounts

XSL•FO
RenderX

have engaged (ie, tweeted at least once) in one of the 40,000 health care terms being tracked, they will be evaluated. The score is not determined by the absolute numbers of tweets or how many mentions they have received for the given time period, but by the impact of the posted messages. The score comprises three components, a social network analysis algorithm, health care stakeholder weighting, and conversation quality algorithm. The network analysis algorithm is inspired by the hyperlink-induced topic search (HITS) algorithm [17] and considers each Twitter account's conversation graph by recursively analyzing the health care influence of each individual conversation partner, the influence of the conversation partner's own conversation partners, and so on [18]. In this respect, it is similar to modern impact factor algorithms for academic journals and Google's PageRank [19]. The score is designed specifically for health care and considers the health care stakeholder groups to which the account holders belong. In other words, it matters what role a person has in health care. If, for example, an account is interacting with or being mentioned by another account that is not related to health care, then those conversations and mentions will have less weight as determined by the algorithm. If, on the other hand, the conversations and mentions are made by a health care stakeholder, then that has more weight according to the algorithm. Based on the analysis of these conversations, a quality score is factored with a conversation volume to provide a weighted measure for the impact scores. After that, the 52 weekly rankings and quality scores are combined into a single number for each social media profile and then normalized on a scale of 0 (very low influence) to 100 (very high influence).

For each of the 1000 Twitter profiles initially included in our stratified random sample, we manually queried their h-indexes in two reference-based databases (Scopus and Google Scholar) by searching their names in each profile service's search engine. Before extracting the data, we used a standardized verification process to confirm if the identified profiles corresponded to the Twitter user. Any profile found in Google Scholar or Scopus was verified using at least three of the following identifiers: name (first, middle, last), title, location (country/city), field/specialty, affiliation, and qualitative analysis of Twitter conversations or a free-form Google search using the associated name and any other identifier available. Once a profile was found in either of the platforms and the verification process confirmed with at least three identifiers, the h-index was extracted. This verification process was created to decrease the probability of extracting data from an incorrect profile (eg, similar name but not the author of interest). In Google Scholar and Scopus, the h-index is calculated as "number *n* of a researcher's papers that have all received at least *n* citations" [20]. Although there have been multiple studies [20-24] that have highlighted the advantages, disadvantages, and variations, no other traditional author-level metric has had the same level of acceptance or resilience over the past 15 years. Individuals can calculate the h-index of any researcher as long as they have access to a resource providing the citation count of that researcher's publications or research objects. The three most prominent resources or platforms that provide citation counts for researchers are the Web of Science (Clarivate Analytics – previously of Thomson Reuters), Scopus (Elsevier), and Google Scholar (Alphabet). To provide a more comprehensive reporting for this study, Scopus and Google Scholar were chosen to provide the traditional/benchmark h-index data for each individual. This decision was based on the 2018 study by Martin-Martin et al [25], which found that the greatest inclusion of citations in Health & Medical Sciences was on Scopus and Google Scholar. Additional factors were considered when choosing between Web of Science and Scopus. Scopus was seen as providing all authors better access to their own author profiles, which would allow authors to clarify their publications and correct inaccuracies. Additionally, a 2016 study by Walker et al [26] showed a higher interrater reliability in Scopus than in Web of Science for the h-index calculation.

The h-index for the first 100 profiles was independently extracted by three independent investigators (LOJS, GM, and TB). In this initial set of 100 profiles, there was a 98% overall agreement for the h-index extracted from Google Scholar and a 96% overall agreement for the h-index extracted from Scopus. Disagreements were discussed and resolved through consensus with the senior author (DC). Once our standardized verification process and data extraction methods exhibited adequate reliability, the remaining profiles were extracted independently; 600 were reviewed by the first author (LOJS) and 150 by each of the two other investigators (GM, TB). Investigators extracting the h-index for these profiles were blinded to the HSG scores of all subjects.

Data Analysis

From the initial list of 1000 subjects, we excluded those with incomplete names or non-individual user profiles. The remaining profiles were included for the main data analysis if an h-index was available from either Google Scholar or Scopus. All analyses were conducted using the BlueSky Statistics (Version 7.0.746.34007) graphic user interface (GUI) for R. Continuous features (HSG score and h-index) from included profiles were summarized as the median and IQR. Correlation analyses were performed between the HSG scores and h-index obtained from Google Scholar (overall h-index and 2015 h-index) and Scopus (overall h-index). Given the highly skewed nature of metrics such as the h-index [27], we calculated the Spearman correlation coefficient (ρ). This is similar to the Pearson correlation, but it is based on ranks rather than original values. Like the Pearson correlation, values range from –1 to +1, with larger absolute values indicating a stronger relationship. A correlation *t* test was conducted to evaluate the statistical significance of the correlation coefficients. *P* values <.05 were considered statistically significant. For sensitivity analysis, we considered the h-index as 0 for those subjects in which a Scopus h-index was not found.

Simple linear regression was initially implemented to understand the linear relationship between the HSG score and h-index (Google Scholar and Scopus). To better understand the true relationship between the HSG score, and the overall h-index provided by Google Scholar and Scopus, negative binomial hurdle regression was performed. The h-index was used as the response variable, and the HSG score was the predictor of interest. A negative binomial model was chosen owing to the skewed nature of the h-index data and the overdispersion present

XSL•FO
RenderX

in the data distribution. To account for the high number of zeroes not covered by a negative binomial distribution, a hurdle model with a binomial logistic link function was also implemented. Model selection was performed using the Vuong test and the Akaike information criteria (AIC).

Results

Twitter Profiles

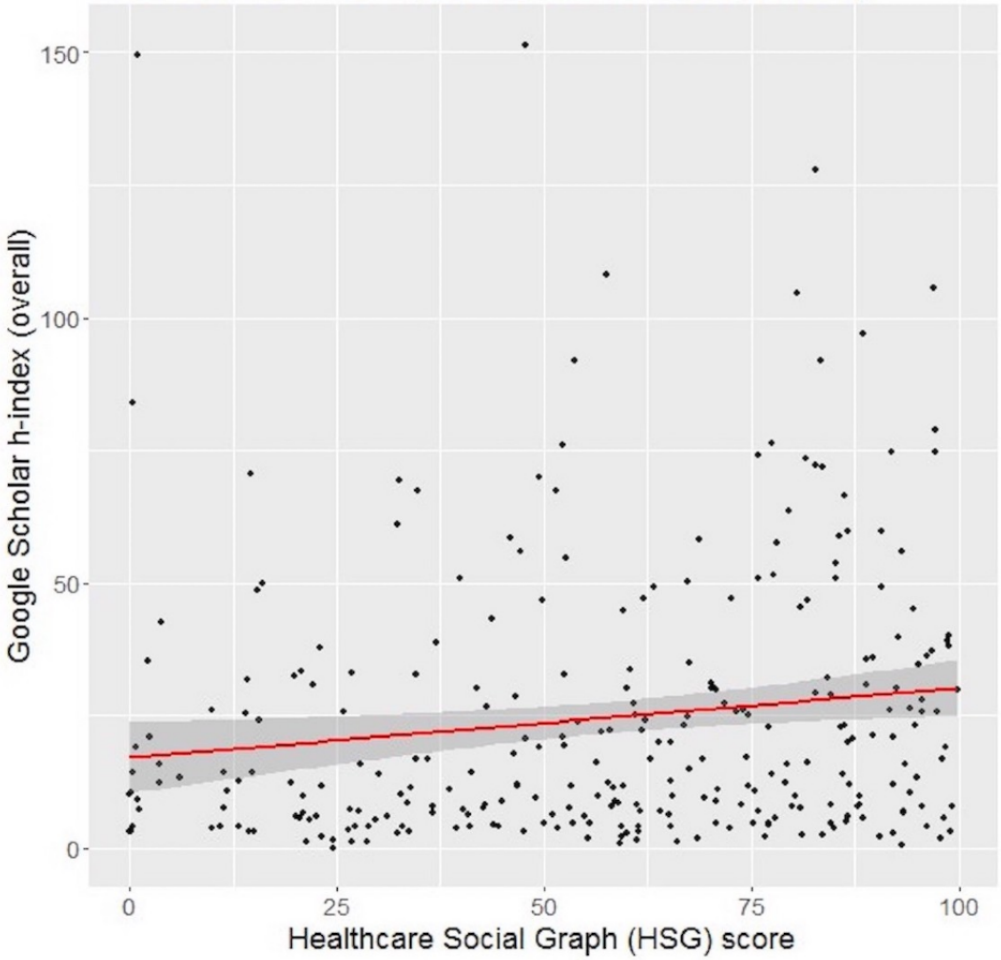
Our stratified random sample generated an initial list of 1000 Twitter profiles from the Symplur database. Of these, 83 were excluded for the following reasons: 5 were repeated profiles, 62 had incomplete names on Twitter (ie, no first and last names, making it impossible to search for a corresponding Google Scholar or Scopus profile), and 16 were not individual user profiles. Among the 917 individual Twitter profiles with complete names for which h-indexes were searchable, 429 (46.8%) were from the United States, 173 (18.9%) from the United Kingdom, 54 (5.9%) from Canada, 49 (5.3%) from Spain, 41 (4.5%) from Australia, 17 (1.9%) from India, 13 (1.4%) from the Netherlands, 13 (1.4%) from France, 12 (1.3%) from Ireland, 9 (1.0%) from Brazil, and the remaining 11.6% from 36 other countries from all continents (only 5 profiles were from unknown countries).

A total of 286 (31.2%) of the 917 stakeholders had a Google Scholar h-index available. The median HSG score for these profiles was 61.1 (IQR 48.2), and the median h-index was 14.5 (IQR 26). A total of 715 (78%) of the 917 stakeholders had a Scopus h-index available. The median HSG score for these profiles was 57.6 (IQR 46.4), and the median h-index was 7 (IQR 16).

Google Scholar h-Index

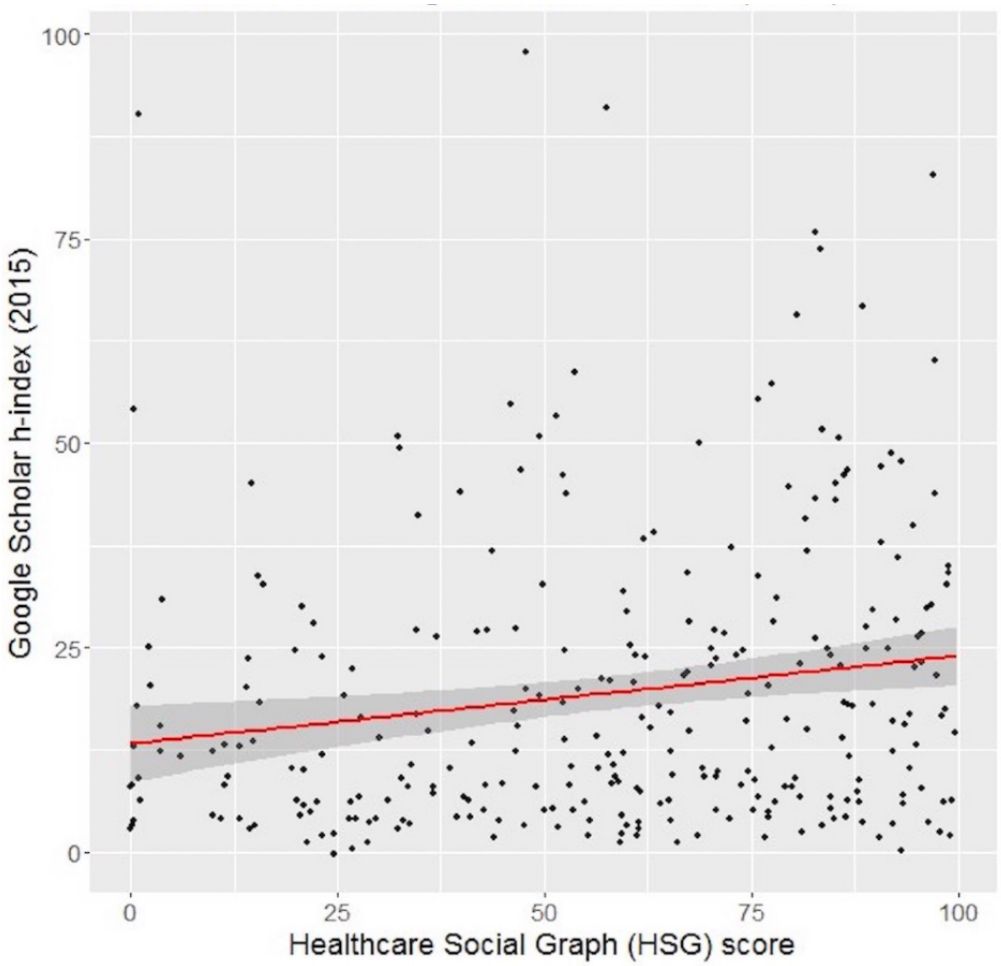
For the 286 subjects with the HSG score and overall h-index provided by Google Scholar available, the Spearman correlation coefficient ρ was 0.1979 (P<.001), indicating a weak positive correlation between these two metrics (Figure 1). When we analyzed the correlation for the 2015 h-index from Google Scholar, the results were similar (ρ=0.203) (Figure 2). Also, when we analyzed the Google Scholar i10 index, the results did not change significantly (see Multimedia Appendix 1 and Multimedia Appendix 2).

Figure 1. Correlation between HSG scores and Google Scholar overall h-indexes. Spearman correlation coefficient ρ=0.1979 (N=286). The red line is the regression line; the shaded area is the 95% CI.



XSL•FO
RenderX

Figure 2. Correlation between HSG scores and Google Scholar 2015 h-indexes. Spearman correlation coefficient ρ=0.203 (N=286). The red line is the regression line; the shaded area is the 95% CI.



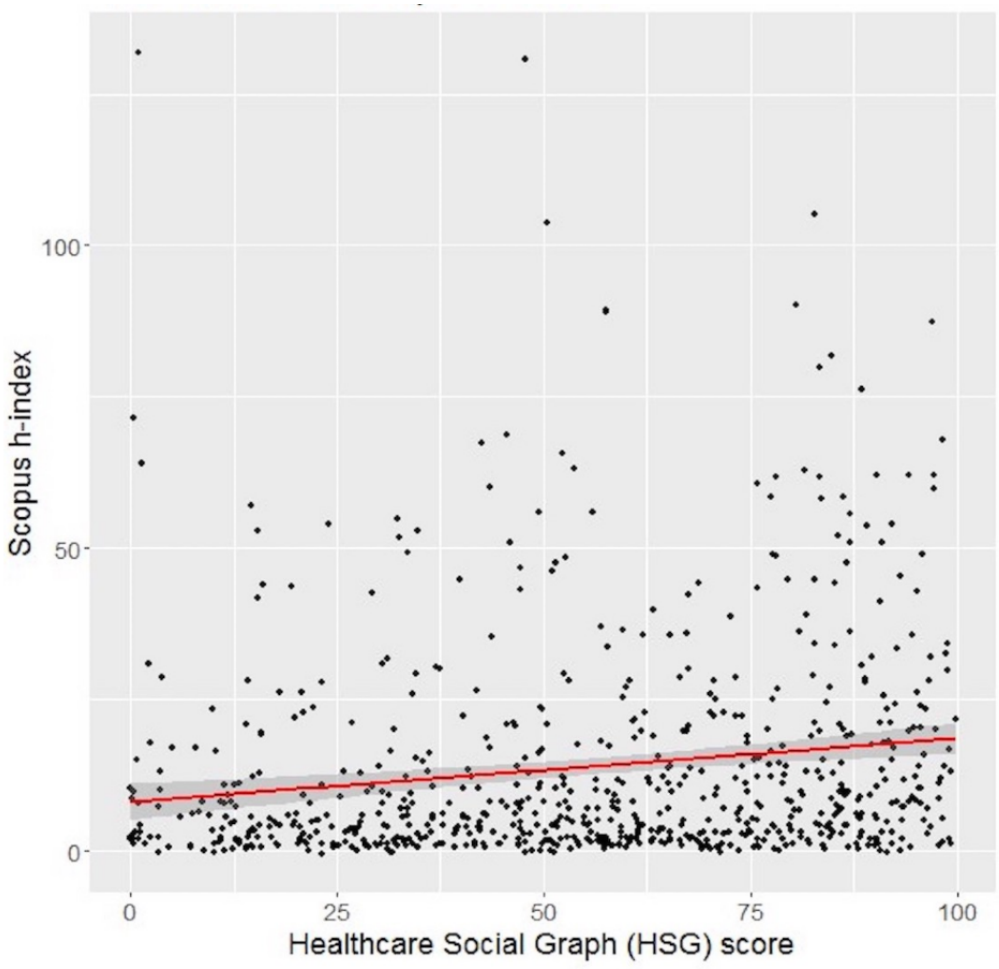
Linear regression using the Google Scholar overall h-index as the response found a significant association between the HSG score and h-index. Assuming a linear relationship, for every 10-point increase in the HSG score, there was an associated increase of 1.134 in the h-index (95% CI 0.280-2.347; *P*=.01). The R^2 value was 0.0214 and the linear regression equation is expressed as (E[Google Scholar overall h-index] = 17.037 + 0.1314*[HSG score]). From the negative binominal hurdle model, there was no effect of the HSG score on whether an author's Google Scholar h-index is 0 or positive (log-odds=0.043; *P*=.31). However, for authors with a positive h-index, a 5-point increase in the HSG score was associated with a 2.7% increase in Google Scholar h-index (exp[coef]=1.027; 95% CI 1.006 -1.048; *P*=.01). Additionally, the Vuong test found that the hurdle model was a better fit than the negative binomial model (*z* statistic=3.092; *P*<.001).

Scopus h-Index

For the 715 subjects with the HSG score and Scopus h-index available, the Spearman correlation coefficient ρ was 0.2173 (*P*<.001), also indicating a weak positive correlation (Figure 3). In the sensitivity analysis, in which subjects without a Scopus h-index available were computed as having an h-index of 0, therefore including all 917 initially eligible profiles, the Spearman correlation coefficient ρ was 0.317 (*P*<.001) (see Multimedia Appendix 3).

XSL•FO
RenderX

Figure 3. Correlation between HSG scores and Scopus h-indexes. Spearman correlation coefficient ρ=0.2173 (N=715). The red line is the regression line; the shaded area is the 95% CI.



Univariate linear regression fitting of Scopus h-indexes found a significant association with the HSG score. Assuming a linear relationship, for every 10-point increase in the HSG, we expect a 1.049-point increase in the h-index (95% CI 0.567-1.530; P<.001). The R^2 value was 0.0249 and the linear regression equation is expressed as (E[Scopus h-index] = 8.0821 + 0.1048*[HSG score]). From the negative binomial hurdle model, we found no significant effect of the HSG score on whether an author's h-index is 0 or positive (log-odds=0.0072; P=.27). However, for authors with a positive h-index, a 5-point increase in the HSG score was associated with a 4% increase in h-index (exp[coef]=1.040; 95% CI 1.021-1.061; P<.001). The Vuong test found that the hurdle model was a better fit than the negative binomial model (z statistic=4.606; P<.001).

Discussion

Principal Results

The advent of digital scholarship is rapidly changing the way scholarship is created and appraised in academia. We are currently seeing a swift transition from a paradigm in which the impact of an academician was circumscribed to deliverables critiqued by a restrictive circle of peers to a novel model in which the importance of scholarly work is measured by the influence and impact it generates in academic circles and among the general public. This leads to the critical need to adopt new appraisal concepts and tools [1,10,28].

The HSG score represents a novel author-level tool within the domain of altmetrics. This metric aims to measure and illustrate the influence a particular stakeholder has in health care social media as a function of user-generated content and interactions. This method is common for analyzing the weight or importance of specific users that are part of larger networks [29]. In general, the more the connections and the more information users create or are involved with, the greater their importance in a network [30].

Citation-focused metrics such as the h-index assess the importance of academicians based on the number of times their work has been cited by other scholars, with a significant bias constructed to value certain outputs (eg, prestigious journals) more than others. This distorts the organic reach and impact of articles (eg, where bad articles in good journals are valued more than good articles in bad journals), conflagrating production and publishability with influence and impact [31].

In this study, we aimed to find if there is a relationship between the HSG score, a marker for influence in a network, and the h-index, a metric for productivity. The assumption driving this comparison was that a high degree of productivity (greater h-index) would be associated with higher impact among

XSL•FO
RenderX

stakeholders in the field and subsequent influence on digital networks.

Approximately three quarters of health care stakeholders identified by HSG scores had a concomitant h-index profile; this simple observation illustrates that there is a significant overlap between academic endeavors and the participation of these users in social media. In other words, academicians are part of general public forums such as social media; they value such forums and participate in them.

When analyzing the relation between the HSG score and h-index, we found a correlation, albeit a weak one, between the two metrics. This positive relation seems to indicate that the higher the HSG score, the higher the h-index (and vice versa). We believe the association describes a relation between scholarly productivity and influence in a health care network; this is possibly explained by academicians using digital domains to disseminate their scholarly work and subsequently bring attention to it, by measuring interactions with other stakeholders and organically increasing their connections and weight in the network. Nevertheless, it is important to mention that the low R^2 value approximately at 2% implies that although we have a statistically significant correlation and are capturing similar trends, there is a sizable amount of variability that is not shared between these two metrics. This emphasizes the simple fact that these metrics measure *different* components of scholarly work and should be evaluated in an independent and complementary way.

The HSG score and h-index are metrics that are of interest for scholars and academic establishments. Per their definitions, these tools are aimed at different aspects of the CRAM framework [8], where the HSG score likely appraises impact and influence and h-index productivity and quality. Remarkably, from our analysis, we can describe an association bridging these four aspects; the influence and impact of a user in a health care–specific digital network are correlated with their academic productivity and quality. More than a chiasm between traditional citation metrics and novel social media–based metrics, our findings point toward a positive relation between the two domains.

Limitations

There are several limitations that need to be acknowledged. First, the accuracy of the metrics that were obtained (which were subsequently used to compute correlations) depends on the validity of the data provided by each reference-based database. Some of these platforms (eg, Google Scholar) can easily be manipulated [32]. Scopus automatically calculates the h-indexes of authors without a profile in their database, which explains why there were higher numbers of profiles and h-indexes available in Scopus when compared to Google Scholar, in which individuals need to create active profiles. In Scopus, authors may have more than one profile and, for this reason, we have used an available tool in their platform to combine profiles from the same author to obtain the most accurate h-index for that author. Second, Scopus and Google Scholar data are dynamic because new citations are constantly being added to their databases. As we were unable to automatically retrieve the h-indexes from these databases on the same day, manual data extraction occurred over a four-month period. Therefore, authors may have had their h-indexes extracted with a time difference as long as 100 days, and this, although unlikely, could have influenced the accuracy of our analysis. We assumed that the h-index would be time invariant (while in fact it is not) during the period of data extraction. Nevertheless, the h-indexes should theoretically be less dynamic than citations alone, and it is unlikely to change by a large magnitude even after a 100-day period [33]. Third, we have not considered the ages of the authors, which might have an impact on the correlation measures given that more experienced authors may exhibit distinctive behavior compared to emerging authors. Fourth, Google Scholar seems to overestimate author-level metrics when compared to Scopus owing to inclusion of gray literature citations, among other reasons. However, we extracted the h-indexes from both databases, and the results did not change when using one h-index over another. In fact, the h-indexes from Google Scholar and Scopus were strongly correlated with each other (see Multimedia Appendix 4).

Conclusions

It appears that novel author-level altmetrics based on network analysis in social media and digital publications have a positive association with traditional bibliometric benchmarks. This seems to indicate that not only can they coexist but can also supplement and augment each other's domains. Academicians interested in a comprehensive appraisal of their academic work and preparing for advancement need to be deliberate about investing time and attention into both spheres of appraisal (traditional and altmetrics), as they are relevant, significant, and most importantly appear to move in the same direction and amplify each other.

Acknowledgments

This research was supported in part through the Center for Clinical and Translational Science (CCaTS) Small Grant Program, part of Mayo Clinic CCaTS grant number UL1TR000135 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH.

Authors' Contributions

LOJS and DC conceived and designed the study. LOJS, GM, and TB conducted the acquisition of the h-index data. AU executed the stratified random sampling from the Symplur database and provided the HSG scores. AM analyzed the data and provided statistical expertise. LOJS, GM, TB, and DC interpreted the data. LOJS and DC drafted the manuscript and all authors contributed

XSL•FO
RenderX

substantially to its revision through critical reviewing of the manuscript for important intellectual content. DC (senior author) was responsible for overseeing the project.

Conflicts of Interest

AU is a Partner of Symplur. AU provided the initial sample of subjects included in the study with their corresponding HSG scores. Data extraction of h-index and data analysis were performed independently by the other investigators. AU did not influence the study design, did not participate in the data extraction of the h-index, and did not participate in the analysis of the results or writing of the conclusions. No other investigators have any conflicts of interest to disclose related to this study.

Multimedia Appendix 1
Correlation between HSG scores and Google Scholar overall i10 indexes. Spearman correlation coefficient ρ=0.2037 (N=286).
[PNG File , 107 KB - jmir_v23i5e28859_app1.png]

Multimedia Appendix 2
Correlation between HSG scores and Google Scholar 2015 i10 indexes. Spearman correlation coefficient ρ=0.2087 (N=286).
[PNG File , 111 KB - jmir_v23i5e28859_app2.png]

Multimedia Appendix 3
Correlation between HSG scores and Scopus h-indexes after computing a Scopus h-index of 0 for subjects without Scopus h-indexes. Spearman correlation coefficient ρ=0.317 (N=917).
[PNG File , 189 KB - jmir_v23i5e28859_app3.png]

Multimedia Appendix 4
Strong correlation between Scopus and Google Scholar h-indexes.
[PNG File , 116 KB - jmir_v23i5e28859_app4.png]

References

1. Cabrera D, Roy D, Chisolm MS. Social media scholarship and alternative metrics for academic promotion and tenure. J Am Coll Radiol 2018 Jan;15(1):135-141. [doi: 10.1016/j.jacr.2017.09.012]
2. Haustein S. In: Haustein S, editor. Multidimensional Journal Evaluation: Analyzing Scientific Periodicals beyond the Impact Factor. Germany: De Gruyter Saur; 2012.
3. Priem J, Piwowar H, Hemminger B. Altmetrics in the wild: Using social media to explore scholarly impact. arXiv Preprint posted online on March 20, 2012. [FREE Full text]
4. Piwowar H. Altmetrics: Value all research products. Nature 2013 Jan 10;493(7431):159-159. [doi: 10.1038/493159a] [Medline: 23302843]
5. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012 Sep 01;19(5):777-781 [FREE Full text] [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
6. The state of altmetrics: a tenth anniversary celebration. Altmetric. URL: https://altmetric.figshare.com/articles/online_resource/The_State_of_Altmetrics_A_tenth_anniversary_celebration/13010000 [accessed 2020-11-16]
7. Priem J, Costello KL. How and why scholars cite on Twitter. Proc. Am. Soc. Info. Sci. Tech 2011 Feb 03;47(1):1-4. [doi: 10.1002/meet.14504701201]
8. Braithwaite J, Herkes J, Churruca K, Long JC, Pomare C, Boyling C, et al. Comprehensive Researcher Achievement Model (CRAM): A framework for measuring researcher achievement, impact and influence derived from a systematic literature review of metrics and models. BMJ Open 2019 Mar 30;9(3):e025320. [doi: 10.1136/bmjopen-2018-025320]
9. About PlumX Metrics. Plum Analytics. URL: https://plumanalytics.com/learn/about-metrics/ [accessed 2020-11-16]
10. 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]
11. San Francisco Declaration on Research Assessment. DORA. URL: https://sfdora.org/read/ [accessed 2021-04-06]
12. Nivash J, Dhinesh BL. An empirical analysis of big scholarly data to find the increase in citations. In: Advances in Intelligent Systems and Computing. Switzerland: Springer; 2019:41-51.
13. Healthcare Social Graph®. Symplur. URL: https://www.symplur.com/technology/healthcare-social-graph/ [accessed 2020-11-16]
14. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg 2014 Dec;12(12):1500-1524 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.014] [Medline: 25046751]
15. Healthcare Stakeholder Segmentation. Symplur. URL: https://help.symplur.com/en/articles/103684-healthcare-stakeholder-segmentation [accessed 2020-02-27]

XSL•FO
RenderX

16. Maharani W, Adiwijaya, Gozali A. Degree centrality and eigenvector centrality in twitter. 2014 Presented at: 2014 8th International Conference on Telecommunication Systems Services and Applications (TSSA); October 23-24, 2014; Bali, Indonesia p. 1-5. [doi: 10.1109/tssa.2014.7065911]
17. Kleinberg JM. Authoritative sources in a hyperlinked environment. J ACM 1999 Sep;46(5):604-632. [doi: 10.1145/324133.324140]
18. Healthcare Social Graph Score. Symplur. URL: https://help.symplur.com/en/articles/103681-healthcare-social-graph-score [accessed 2020-02-27]
19. PageRank. Wikipedia. URL: https://en.wikipedia.org/wiki/PageRank [accessed 2020-02-27]
20. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572 [FREE Full text] [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
21. Bornmann L, Daniel HD. What do we know about the h index? J Am Soc Inf Sci Technol 2007 Jul;58(9):1381-1385. [doi: 10.1002/asi.20609]
22. Costas R, Bordons M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. J Informetr 2007 Jul;1(3):193-203. [doi: 10.1016/j.joi.2007.02.001]
23. Teixeira da Silva JA, Dobránszki J. Multiple versions of the h-index: Cautionary use for formal academic purposes. Scientometrics 2018 Feb 20;115(2):1107-1113. [doi: 10.1007/s11192-018-2680-3]
24. Costas R, Franssen T. Reflections around 'the cautionary use' of the h-index: Response to Teixeira da Silva and Dobránszki. Scientometrics 2018 Mar 9;115(2):1125-1130 [FREE Full text] [doi: 10.1007/s11192-018-2683-0] [Medline: 29628537]
25. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado López-Cózar E. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. J Informetr 2018 Nov;12(4):1160-1177. [doi: 10.1016/j.joi.2018.09.002]
26. Walker B, Alavifard S, Roberts S, Lanes A, Ramsay T, Boet S. Inter-rater reliability of h-index scores calculated by Web of Science and Scopus for clinical epidemiology scientists. Health Info Libr J 2016 Jun 11;33(2):140-149 [FREE Full text] [doi: 10.1111/hir.12140] [Medline: 27168256]
27. de Solla Price DJ. Networks of scientific papers. Science 1965 Jul 30;149(3683):510-515. [doi: 10.1126/science.149.3683.510] [Medline: 14325149]
28. Husain A, Repanshek Z, Singh M, Ankel F, Beck-Esmay J, Cabrera D, et al. Consensus guidelines for digital scholarship in academic promotion. West J Emerg Med 2020 Jul 08;21(4):883-891 [FREE Full text] [doi: 10.5811/westjem.2020.4.46441] [Medline: 32726260]
29. Riddell J, Brown A, Kovic I, Jauregui J. Who are the most influential emergency physicians on Twitter? West J Emerg Med 2017 Feb 01;18(2):281-287 [FREE Full text] [doi: 10.5811/westjem.2016.11.31299] [Medline: 28210365]
30. Barabási AL. Scale-free networks: A decade and beyond. Science 2009 Jul 24;325(5939):412-413. [doi: 10.1126/science.1173299] [Medline: 19628854]
31. Waltman L, van Eck NJ. The inconsistency of the h-index. J. Am. Soc. Inf. Sci 2011 Oct 31;63(2):406-415. [doi: 10.1002/asi.21678]
32. Delgado López-Cózar E, Robinson-García N, Torres-Salinas D. The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. J Assn Inf Sci Tec 2013 Nov 11;65(3):446-454. [doi: 10.1002/asi.23056]
33. Schreiber M. How relevant is the predictive power of the h-index? A case study of the time-dependent Hirsch index. J Informetr 2013 Apr;7(2):325-329. [doi: 10.1016/j.joi.2013.01.001]

Abbreviations

AIC: Akaike information criteria
API: application programming interface
CRAM: comprehensive researcher achievement model
DORA: Declaration on Research Assessment
GUI: graphical user interface
HITS: hyperlink-induced topic search
HSG: Healthcare Social Graph
IQR: interquartile range
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by R Kukafka; submitted 16.03.21; peer-reviewed by T Chan, M Janodia; comments to author 05.04.21; revised version received 07.04.21; accepted 06.05.21; published 31.05.21.

Please cite as:
Oliveira J e Silva L, Maldonado G, Brigham T, Mullan AF, Utengen A, Cabrera D
Evaluating Scholars' Impact and Influence: Cross-sectional Study of the Correlation Between a Novel Social Media–Based Score and an Author-Level Citation Metric
J Med Internet Res 2021;23(5):e28859
URL: https://www.jmir.org/2021/5/e28859
doi:10.2196/28859
PMID:34057413

©Lucas Oliveira J e Silva, Graciela Maldonado, Tara Brigham, Aidan F Mullan, Audun Utengen, Daniel Cabrera. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2021. 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

Leveraging Genomic Associations in Precision Digital Care for Weight Loss: Cohort Study

Ranjan Sinha[1], MBA; Dashyanng Kachru[1,2], MS; Roshni Ray Ricchetti[1]; Simitha Singh-Rambiritch[1], BDS, MSc; Karthik Marimuthu Muthukumar[1], MS; Vidhya Singaravel[1], BE, MBA; Carmel Irudayanathan[1], BS, MHRM; Chandana Reddy-Sinha[1], BS; Imran Junaid[1], MS, MD; Garima Sharma[1], MBBS; Patricia Alice Francis-Lyon[1,2], MS, PhD

[1]Digbi Health, Los Altos, CA, United States

[2]Health Informatics, University of San Francisco, San Francisco, CA, United States

Corresponding Author:
Patricia Alice Francis-Lyon, MS, PhD
Digbi Health
13105 Delson Ct
Los Altos, CA, 94022
United States
Phone: 1 510 883 3721
Email: patricia@digbihealth.com

Abstract

Background: The COVID-19 pandemic has highlighted the urgency of addressing an epidemic of obesity and associated inflammatory illnesses. Previous studies have demonstrated that interactions between single-nucleotide polymorphisms (SNPs) and lifestyle interventions such as food and exercise may vary metabolic outcomes, contributing to obesity. However, there is a paucity of research relating outcomes from digital therapeutics to the inclusion of genetic data in care interventions.

Objective: This study aims to describe and model the weight loss of participants enrolled in a precision digital weight loss program informed by the machine learning analysis of their data, including genomic data. It was hypothesized that weight loss models would exhibit a better fit when incorporating genomic data versus demographic and engagement variables alone.

Methods: A cohort of 393 participants enrolled in Digbi Health's personalized digital care program for 120 days was analyzed retrospectively. The care protocol used participant data to inform precision coaching by mobile app and personal coach. Linear regression models were fit of weight loss (pounds lost and percentage lost) as a function of demographic and behavioral engagement variables. Genomic-enhanced models were built by adding 197 SNPs from participant genomic data as predictors and refitted using Lasso regression on SNPs for variable selection. Success or failure logistic regression models were also fit with and without genomic data.

Results: Overall, 72.0% (n=283) of the 393 participants in this cohort lost weight, whereas 17.3% (n=68) maintained stable weight. A total of 142 participants lost 5% bodyweight within 120 days. Models described the impact of demographic and clinical factors, behavioral engagement, and genomic risk on weight loss. Incorporating genomic predictors improved the mean squared error of weight loss models (pounds lost and percent) from 70 to 60 and 16 to 13, respectively. The logistic model improved the pseudo R^2 value from 0.193 to 0.285. Gender, engagement, and specific SNPs were significantly associated with weight loss. SNPs within genes involved in metabolic pathways processing food and regulating fat storage were associated with weight loss in this cohort: rs17300539_G (insulin resistance and monounsaturated fat metabolism), rs2016520_C (BMI, waist circumference, and cholesterol metabolism), and rs4074995_A (calcium-potassium transport and serum calcium levels). The models described greater average weight loss for participants with more risk alleles. Notably, coaching for dietary modification was personalized to these genetic risks.

Conclusions: Including genomic information when modeling outcomes of a digital precision weight loss program greatly enhanced the model accuracy. Interpretable weight loss models indicated the efficacy of coaching informed by participants' genomic risk, accompanied by active engagement of participants in their own success. Although large-scale validation is needed, our study preliminarily supports precision dietary interventions for weight loss using genetic risk, with digitally delivered recommendations alongside health coaching to improve intervention efficacy.

(J Med Internet Res 2021;23(5):e25401) doi:10.2196/25401

XSL•FO
RenderX

KEYWORDS

obesity; digital therapeutics; precision nutrition; nutrigenomics; personalized nutrition; mHealth; mobile apps; gut microbiota; machine learning; health coaching; lifestyle medicine; mobile phone

Introduction

Background

The global death toll of COVID-19 has eclipsed 1 million cases [1]. Obesity, following age, has emerged as the most critical risk factor for morbidity, hospitalizations, and complications [2]. The prevalence of obesity in the United States and in other Western countries has increased sharply in the last 2 decades. Since the early 1960s, when more than 10% of Americans were obese, that proportion has grown to 42.4% of adults [3]. Moreover, the prevalence of obesity is higher in minority communities: 49.6% of non-Hispanic Black individuals and 44.8% of Hispanic Americans are obese, compared with 42.2% of non-Hispanic White individuals. These same minority communities are experiencing disproportionate COVID-19–driven mortality, likely linked, at least in part, to the heightened prevalence of obesity [4]. Although the precise cause of obesity is yet to be discovered, several factors have been linked to its development [5]. In particular, biology interacts with behavior and demographics (such as socioeconomic status or ethnic and cultural cuisine) to influence obesity risk [6]. Obesity-associated biological factors include, but are far from limited to, genetics and epigenetics, microbiome composition, age, circadian rhythm disruption, pharmaceutical interactions, and comorbidities and their management [6,7].

The rapid increase in obesity prevalence has coincided with sociological factors, such as generally reduced physical activity alongside a rise in the consumption of highly processed, high-calorie, but nutrient-poor foodstuffs. However, these obesogenic conditions did not affect the population uniformly. Instead, a notable proportion of the population is still able to remain at a healthy weight, indicating that the heterogeneous response to obesogenic conditions may result, in part, from individual innate protection from these conditions, possibly conferred by the genetic makeup [8].

Most current clinical interventions for obesity management focus on lifestyle and dietary adaptation with varying levels of professional guidance and involvement, short- or long-term pharmaceutical therapies, and bariatric surgery [9]. Individual responses to these therapeutic interventions are confoundingly (for clinicians and participants alike) heterogeneous for multifactorial reasons [10], making the need for personalized, precision medicine courses of treatment imperative. Most Americans (63%) have made serious efforts toward weight loss over the course of their lives, and almost one-third are trying to lose weight [11]. In 2014, commercial weight loss services were a US $2.5 billion market consisting primarily of the following market shares—Weight Watchers (45%), NutriSystem (14%), and Jenny Craig (13%)—but the long-term effectiveness of various commercial *calorie restriction*–based weight loss programs is unclear [12-14] (Table 1).

Table 1. Commercial weight loss services

Program	Market share (%)	Intervention	Cost per month, US $
Jenny Craig	13	Low-calorie meal replacements with one-on-one counseling	>450
Nutrisystem	14	Low-calorie meal replacements with one-on-one counseling	300-350
Weight Watchers	45	Self-monitoring with web-based coaching and points tracking	43 (plus food)

Personalizing Weight Loss Interventions

Recent research has elucidated the mechanisms of food-derived biomarkers, allowing for stratification based on a participant's unique metabolism of given food products. This permits the targeting of personalized nutrition to groups that are better characterized [8,15,16]. For example, given that low-grade inflammation has been implicated in insulin resistance, mediating inflammation via targeted dietary approaches is a precision nutrition intervention [17,18].

Advances have already been made in the early intervention and risk assessment of participants who are obese by designing therapies based on unique genetic predisposition and risk. Environmental interventions such as diet and exercise can trigger epigenetic changes, altering gene expression in metabolic pathways. Recent research indicates that physical activity and high-fat diets may alter DNA methylation patterns in skeletal muscle and adipose tissue [19-21], influencing weight management [8]. Eventually, researchers hope to elucidate the genetic patterns that influence individual obesity and concomitant illness susceptibility, risk of progression, and response to therapy, to provide participants with optimal treatment [22].

Epigenetics and Their Role in Obesity

Even as science illuminates many genetic risk factors of complex metabolic diseases such as obesity and type 2 diabetes [23-27], these genetic variants account for only a fraction of BMI variation [25]. The *missing* heritability of obesity might be at least partially explained by interactions between genetics and environmental factors [28]. In particular, specific gene variants may influence sensitivity to certain environmental factors so that exposure to these factors in susceptible individuals can contribute to disease. As individuals who are obese are characterized by considerable heterogeneity within the spectrum of clinical obesity, supporting gene-diet interaction and precision nutrition in different subtypes of obesity is imperative [29-33].

Bariatric surgery is a weight loss option for participants with severe and complex obesity, for whom dietary interventions or digital therapeutics have been less than successful [34-36].

XSL•FO
RenderX

Genetics may be a significant predictor of weight loss following Roux-en-Y gastric bypass surgery [37], but few genetic variants have been characterized to date [38,39].

The Role of Diet in Obesity

Although obesity can, in some cases, be linked to excessive appetite and food consumption, these behaviors may have a genetic component, and even food preferences themselves may have a genetic basis [40,41]. For example, the alpha-ketoglutarate-dependent dioxygenase (FTO) locus rs9939609 has been associated with reduced satiety [42], increased caloric and fat intake [43,44], and a propensity to consume calorie-dense foods [43,45]. The TAS2R38 genotype differentiates potential super-, medium-, and nontasters of bitter-tasting thiourea compounds. These different bitter-tasting profiles appear to be predictive of differential dietary preferences, and in particular, nontasters were observed to have higher BMIs [46]. Considered together and alongside other evidence, this research implies that body weight and BMI may be affected by genetic variations in food preferences, tendencies, and eating behaviors. Elucidating how food intake and body metrics are mediated by genetics is challenging because of the difficulty of reproducing results across varying populations and the complexity of identifying causal interactions [47-49]. Research using randomized controlled trials (RCTs) and large sample–sized biobanks with electronic health records will better characterize how diet and genetics interact to mediate health outcomes [50,51].

The Role of Physical Activity in Obesity

Exercise that can prevent weight gain and promote weight maintenance has been well established through research [52-55]. Evidence suggests that body weight, waist-to-hip ratio, and BMI are significantly associated with adherence to an aerobic exercise intervention [56]. Interestingly, the propensity for exercise appears to be heritable, at least in part, with studies estimating heritability ranging from 9% to almost 80% [57]. MC4R genes appear to be associated with physical inactivity [58]; however, other genes may be associated with adherence and tolerance to physical activity regimens [56].

Gut Microbiome and Its Role in Obesity

The human gastrointestinal tract hosts millions of commensal micro-organisms comprising the gut microbiome, which acts as a virtual endocrine organ regulating nutrient production and metabolism, satiety, and even energy homeostasis [8,59]. These microbes are intrinsically linked to host health, as they are implicated in nutrient processing and metabolism, pathogen displacement, vitamin synthesis, and body weight regulation [60]. Researchers and clinicians have been studying alterations of the gut microbiome in individuals, as perturbations in the gut microbiome appear to underlie the pathophysiology of obesity and associated comorbidities, such as type 2 diabetes and metabolic syndrome [61,62]. Microbiome profiling for nutritional intervention is gaining prominence as a key feature of precision nutrition.

Research on the impact of specific dietary factors on microbiome diversity can guide interventions focused on optimizing gut microbial composition [63]. For example, variation in the lactase (LCT) gene region, associated with response to dairy intake, appears to be associated with the abundance of the gut microbiome Bifidobacterium [64]. In particular, variations in LCT were found to be predictive of the obesity-based modulation of dairy lactose and milk intake [65], indicating that shifts in gut microbiota across LCT genotypes could be tied to the caloric extraction of ingested food [65]. Similar to specific genes, specific bacterial species have also been directly implicated in the etiology of obesity. Methanobrevibacter smithii, for example, can itself metabolize dietary substrates or metabolic byproducts of other bacteria, thereby promoting weight gain [66].

Further evidence ties both an individual's genetics and diet to microbiome composition because lower microbial diversity appears to be associated with excess weight gain [67]. Even in early childhood, disruptions in the gut microbiome can have a long-lasting influence on adult body weight [68]. Moreover, nutritional interventions such as administering prebiotics and probiotics to manipulate gut microbiota that promote or are refractory to weight loss show potential as obesity interventions but require further study [69]. Weight loss, whether mediated by diet or via bariatric surgery, can alter the gut microbiome in ways that affect the efficacy of various weight loss strategies [70,71]. An interesting feature of bariatric surgery is that it appears to induce obesity-associated gut microbiota to shift toward lean microbiome phenotypes [72].

Behavioral and Digital Interventions in Obesity

As the obesity epidemic continues to proliferate, new digital programs available on websites or as smartphone apps are being leveraged to promote weight loss [73]. Digital programs are agile in that they can easily be modified to reflect the latest research and best practices in a rapidly changing field; they are more cost-effective than traditional in-person programs and are also more easily scalable, increasing their reach [74,75]. Resources can include activity trackers, videos, logs, device-to-device communication, and third-party app compatibility [76]. In addition, research indicates that remotely administered programs can result in significant weight loss [76-79].

Digital programs have the ability to provide personalization to address the plethora of needs presented by participants [76]. Individuals partaking in such programs are still able to leverage interpersonal relationships. Digital health coaching, for example, allows participants to discuss their weight loss journey via any number of communication platforms [75]. According to research, both in-person and telehealth coaching relationships are effective in motivating overweight individuals to work toward weight loss [80]. In a recent study of more than 600 participants in a smartphone-based weight loss program with a coaching component, participants lost, on average, more than 7% of their body weight, successfully passing the 5% weight loss marker that many in-person programs set [75].

The multifactorial nature of obesity is reflected in the myriad heritable, behavioral, and environmental factors that can lead to obesity risk [47]. The most successful interventions are likely to be those that leverage current findings across the full spectrum of obesity-related risk factors: dietary interventions accounting

XSL•FO
RenderX

for both genetic markers of food sensitivity, metabolic predispositions, and behavioral risk as well as those geared toward optimizing gut microbial diversity and composition; physical activity measures taken in consideration of genetic risk profiles; and behavioral modifications undertaken via digital care [59]. The precision nutrition therapy offered by Digbi Health aims to account for these various factors in delivering a personalized course of obesity intervention [59]. In this study, we describe the weight loss of a cohort of participants enrolled in the Digbi Health program, modeling and analyzing genomic, lifestyle and engagement factors that were found to be influential in this cohort. It was hypothesized that weight loss models would exhibit a better fit when incorporating genomic data than when using demographic and engagement variables alone.

Methods

Recruitment

For this study, we identified all Digbi Health participants who were enrolled between June 2019 and June 2020, had been in the program for at least 120 days, and had been genotyped by Digbi Health. Sample collection kits were shipped to 443 participants, of which 393 mailed back their samples for processing, thereby yielding a cohort size of 393. Among these participants, 315 individuals self-identified as female, 77 individuals self-identified as male, and one individual declined to state. All participants self-enrolled for the Digbi Health therapy via a large California-based insurance payor wellness program. The qualifying criteria to join the program were BMI >25 kg/m^2 with a comorbidity (eg, prediabetes, diabetes, cardiovascular disease, or hypertension) or BMI >30 kg/m^2, regardless of comorbidities. Participants were advised to remain under the care and supervision of their existing physicians and were further advised to notify physicians and other health care providers of their participation in the Digbi Health program. The data set included data from each participant's first 120 days in the program. This Digbi Health anonymized, retrospective research study was exempted from full review by the Ethical and Independent Review Services West Coast Board, Corte Madera, California, reference 20149-01. All participants agreed to the Digbi Health terms and conditions and privacy policy when enrolling in the therapy.

Intervention

Digbi Health is a next-generation, prescription-grade, digital therapeutic platform that uses artificial intelligence to analyze genetics, gut bacteria, lifestyle habits, and socioeconomic and behavioral risk patterns to create evidence-based personalized nutrition, fitness, sleep, and stress management programs to reduce weight and reverse weight-related inflammatory gut, musculoskeletal, cardiovascular, and insulin-related diseases. Digital precision care interventions are delivered via web-based or mobile apps to expand the accessibility, safety, and effectiveness of health care. Digbi Health's individualized program is geared primarily toward individuals who are overweight or obese, with or without a comorbidity, and functions as a weight loss management tool. The therapy is currently covered by a large California-based health insurance payor for their qualifying members through its obesity management wellness platform.

On enrolling in the Digbi Health program, participants were provided with web-based log-in credentials and were mailed a Bluetooth-compatible digital weighing scale and saliva and stool biosampling kits. App usage consisted of daily tracking of weight (via the Bluetooth scale), tracking of dietary intake (uploading photographs of all food items consumed), and tracking wellness-associated metrics (sleep quality and quantity, exercise type and duration, stress and meditation, energy levels, cravings, and recommended foods consumed or avoided).

Sample Collection

The individual's DNA was self-collected using a buccal swab (Mawi Technologies iSwab DNA collection kit, model no. ISWAB-DNA-1200). Saliva DNA extraction, purification, and genotyping using Affymetrix Direct to Consumer Array version 2.0 on the Affymetrix GeneTitan was all performed by the AKESOgen laboratory. The results presented in the genetics section of the report were determined by the number of markers and risk genotypes present in the genomic raw data, the Digbi Health reports were loaded into the app, and coaching was individualized based on participants' genomic risk factors. Individuals' gut microbiomes were self-collected via a fecal swab (Mawi Technologies iSWAB Microbiome collection kit, model no. ISWAB-MBF-1200). Sample processing and 16SrRNA-targeted next-generation sequencing were performed at the AKESOgen laboratory. Although the app and coaching are personalized based on participants' microbiome data, these data were not analyzed in this study and are the subject of a forthcoming research article. The personalized Digbi Health plan was systematically reviewed with the participants in individualized sessions with the health coach over a 4-month period at regular, predetermined, weekly, and biweekly intervals.

Genetic Report

The Digbi Health genetic report consisted of two sections: gene nutrition and gene fitness. The gene nutrition report analyzed participants' genotypes that have been shown to influence nutritional traits, such as diet and weight management, micronutrient requirements, food intolerance and sensitivity, and several other attributes relevant to nutritional well-being. For each of these traits, participants were assigned a *high*-, *medium*-, or *low*-risk score based on the number of risk alleles detected, and health coaches guided interventions based on these potential risks (eg, suggesting someone with high risk for gluten intolerance eliminates dietary gluten or someone with medium risk reduces consumption). The degree of risk associated with any specific single-nucleotide polymorphism (SNP) was determined by the presence of 0, 1, or 2 risk alleles. Several individual SNPs may have contributed to a single trait or function, and some of these SNPs might have increased the risk for a trait, whereas others may have decreased it. In Digbi Health gene reports, as many SNPs as possible were considered when determining the risk of a particular trait. Although Digbi Health coaching is individualized based on several different traits, a number of notable traits and associated SNPs and how risk factors for these traits inform individualized health coaching

XSL•FO
RenderX

have been highlighted in the *Results* and *Discussion* sections of this paper.

The gene fitness report analyzed SNPs studied in conjunction with fitness regimes, exercise motivation, and the ability to develop various types of muscle fibers. This section of the report also analyzed the potential inflammatory response to exercise, including endurance, strength, and flexibility training. As in the gene nutrition section, each trait was assigned a *high*-, *medium*-, or *low*-risk score based on SNP data, and health coaches guided participants through recommendations for healthy exercise.

Gut Microbiome Report

In addition to using genetic risk profiles to guide the course of participants' precision care, the Digbi Health program also analyzed gut microbiome profiles (collected from stool swab sampling) to guide the course of care. However, in this study, we aimed to analyze only the effect of demographics and lifestyle and genomic factors on weight loss; the incorporation of the related microbiome data and how they inform individualized health coaching is the subject of forthcoming analyses from our group, currently in preparation.

Lifestyle

The Digbi Health therapy is a 120-day program that uses body metrics, gut microbiome and genetic profiles, and personalized health coaching to manage weight loss. Participants use the Digbi Health app to track 10 key lifestyle and wellness markers (weight, sleep, hunger, cravings, stress, meditation, superfoods, morning energy, foods to avoid, and exercise) on a daily basis and take photos of the food they consume. Each participant is assigned a health coach who works personally with the participant through 12 guided sessions at various intervals to interpret the personalized wellness reports generated from sampling participants' DNA and gut microbiota. The reports also provide a breakdown of obesity risk based on individuals' genetic and gut microbiome profiles. The program is geared toward participants losing at least 5% of their baseline body weight by day 90 of the 120-day program. To achieve this goal, the program seeks to nudge participants toward making incremental lifestyle changes focused on reducing sugar consumption, timing meals to optimize insulin sensitivity, reducing systemic inflammation by identifying possible inflammatory and anti-inflammatory nutrients via genetic testing, and establishing a base level of physical activity. The personalized incremental behavioral modifications suggested by the program are designed to reduce inflammation, optimize gut health based on microbiome testing, and most importantly are supported by health coaching and the app to integrate into the participant's lifestyle so as to be sustainable long term. The genetic profile of Digbi Health users identifies several nutrient and food risk factors that have associations with obesity, comorbidity, or inflammatory risk (eg, gluten sensitivity, lactose tolerance, caffeine sensitivity, fatty acid metabolism, blood pressure response to salt or riboflavin intake, or reduced insulin resistance with exercise), and health coaching guidance is tailored specifically to incorporate participants' risk profiles.

Statistical Analysis

The data from our cohort of 393 participants over their first 120 days in the Digbi Health personalized digital weight loss program were analyzed retrospectively. Interpretable regression models (linear and logistic) were built, and visualizations generated using R software (R Core Team). Modeling of demographic and behavioral engagement was conducted by fitting 2 linear regression models of weight loss (pounds lost and percentage lost) in this cohort as a function of the variables listed in Table 2.

Table 2. Mean demographic and engagement variables overall and by gender.

Variables	Values, mean (SD)	Males, mean (SD)	Female, mean (SD)
Starting BMI	34.77 (6.66)	33.75 (5.36)	35.01 (6.92)
Age (years)	45.06 (12.02)	46.82 (11.58)	44.63 (12.1)
Number of weight entries	146.39 (133.01)	142.62 (110.18)	147.31 (138.16)
Number of food photos posted	115.2 (139.33)	108.64 (115.59)	116.81 (144.65)
Number of coaching sessions completed	4.95 (2.99)	5.17 (2.98)	4.9 (3)

Genomic-enhanced models were built by incorporating 197 SNPs from participant genomic data as predictors, using Lasso regression on SNPs for variable selection, and then fitting a model to the data set after adding the selected SNPs to the previous engagement variables. The 197 genomic variables were from Digbi-curated panels of SNPs associated with obesity, fitness, nutrient metabolism, and inflammatory markers (Table S1 in Multimedia Appendix 1 [81-88]). Each SNP value was encoded for each participant as their number of risk alleles (0, 1, or 2). One participant did not identify gender, therefore was excluded from all models, resulting in 392 observations included in each of the 4 linear regression models.

Success or failure logistic regression models were also fit, with and without genomic data. Genomic variables were similarly selected from the full panel of 197 SNPs using Lasso logistic regression. Success was defined as ≥5% weight loss, failure as weight gain or negligible change of <2 lb (0.9 kg). Removed from this model were observations of participants who were only partially successful, having lost weight but without reaching the milestone of 5% weight loss. This resulted in the inclusion of 251 cohort participants in the logistic models, both genomic-enhanced and demographic and engagement only.

Insignificant variables were removed from each model, resulting in 6 final interpretable models, half containing demographic and behavioral engagement variables only, whereas the remaining 3 were genomic enhanced.

XSL•FO
RenderX

Demographic variables included gender, age, and baseline BMI. Behavioral engagement variables included the number of coaching sessions completed, number of weight entries, and number of food posts (Table 2). As the number of food posts and number of weight entries were highly correlated (Pearson correlation 0.98), each regression model could include one but not both. To incorporate both variables in modeling, the number of food posts was retained as a predictor for the linear models, whereas the number of weight entries was kept as a predictor for the logistic models.

For genomic-enhanced models, SNP variables were imputed to the most frequent value (mode). SNPs with >10% missing information, high (≥0.8) Pearson correlation with another variable, or zero variance were removed, resulting in 124 SNPs remaining for linear and 122 SNPs remaining for logistic model variable selection by Lasso regression. The SNPs with nonzero coefficients after Lasso regularization for that particular outcome variable (pounds lost, percentage weight loss, and successful weight loss) then served as predictors, along with the three demographic variables and two engagement variables (number of coaching sessions completed along with either number of weight entries or number of food posts).

Results

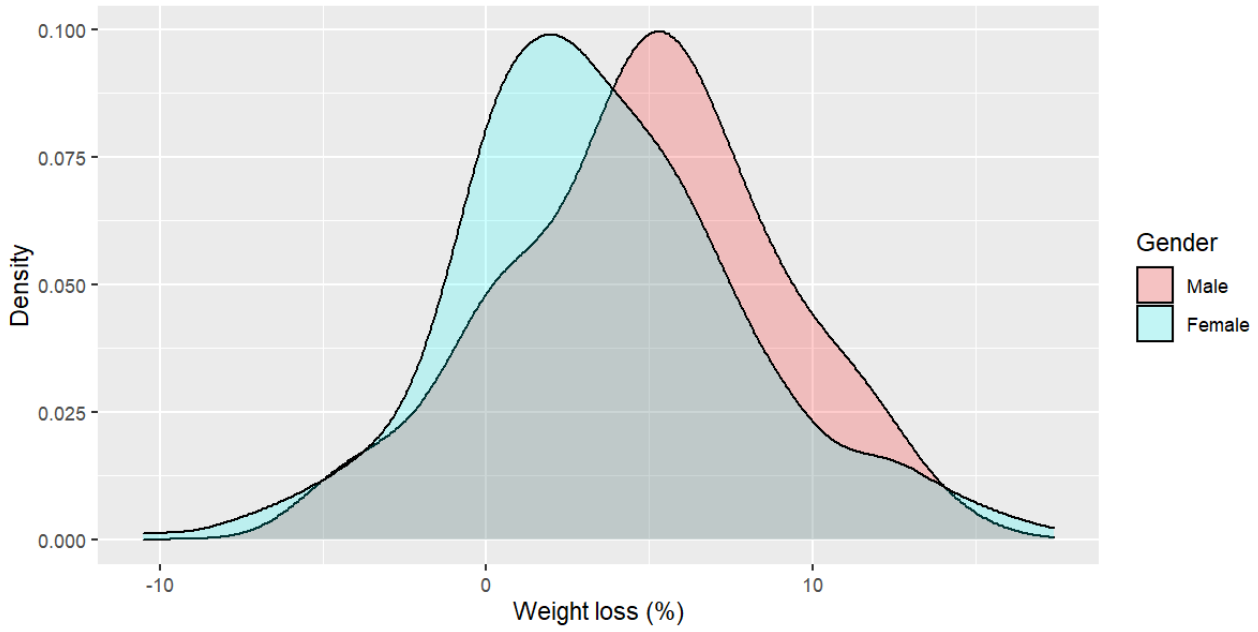
Weight Loss

A total of 393 participants were included in this study to describe and model the weight loss of participants enrolled in the Digbi Health program for 120 days. Of these, 80.4% (315/392) were female and 19,6% (77/392) were male, and one participant declined to state. Tables S1 and S2 in Multimedia Appendix 1 provide a full distribution of the baseline variables. A total of 72% (283/393) participants lost weight compared with 10.7% (42/393) who gained weight, whereas for 17.3% (68/393) participants, the weight remained within normal fluctuations (Table S3, Multimedia Appendix 1). A total of 142 participants lost ≥5% of their baseline body weight within the first 120 days. Improvement in both BMI measures and BMI class over 120 days of treatment is evident in Figure S1, with 25.0% of participants having lost enough weight to move to a lower BMI class (Table S4 in Multimedia Appendix 1). BMI class was defined as presented in Table S5 of Multimedia Appendix 1. The distribution of engagement variables, overall and by gender, is presented in Table 2. In our cohort, no significant difference was found in these variables between males and females (Welch two-sided two-sample *t* tests; results not shown). End points by the obesity class are presented in Table S4 of Multimedia Appendix 1.

As hypothesized, the addition of genomic predictors substantially improved the fit of weight loss models. For linear regression weight loss models (pounds lost and percent), the addition of genomic data improved the mean squared error from 70 to 60 and 16 to 13, respectively, whereas the logistic success or fail model improved pseudo R^2 from 0.193 to 0.285.

Figure 1 depicts the distribution by gender of percent weight loss. The difference in percent weight loss for males and females was found to be statistically significant (Welch two-sided two-sample *t* test, *P*=.02). At an average of 4.8 (SD 4.2) percent of body weight lost, the difference in weight loss between males and females was 1.3 (SD 0.5) percent of body weight. Gender was significant to all linear regression models (Tables S6-S9 in Multimedia Appendix 1) but not to the logistic success or fail models (Tables S10 and S11 in Multimedia Appendix 1), as both women and men succeeded in 5% weight loss within 120 days.

Figure 1. Weight loss (%) distribution by gender.



XSL•FO
RenderX

Significant Variables

Unsurprisingly, baseline BMI was significant to both pounds lost linear models (Tables S6 and S7 in Multimedia Appendix 1) but not for any other model. The participant's age was not significant in any of the models. Increased completion of coaching sessions was significantly associated with increased weight loss in all regression models (Tables S6 to S11 in Multimedia Appendix 1). The two highly correlated engagement variables, number of weight entries and number of food posts, were significant to all models in which they were considered (as described earlier, weight entries were in logistic models, whereas food posts were in linear models; Tables S6 to S11 in Multimedia Appendix 1).

Significant SNPs

In addition to the demographic and engagement variables described earlier, the genomic-enhanced models identified 10 SNPs that were significant to the linear pounds lost model (Table S7 in Multimedia Appendix 1), 11 SNPs to the linear weight loss percentage model (Table S9 in Multimedia Appendix 1), and 6 SNPs to the logistic model (Table S11 in Multimedia Appendix 1). Of the SNPs found significant to the linear models, 8 SNPs were common in both genomic-enhanced linear models (Tables S7 and S9 in Multimedia Appendix 1). In total, 3 notable SNPs that were found to be strongly associated with changes in body weight for this cohort, rs17300539_G, rs2016520_C, and rs4074995_A, were further explored. The literature suggests explanatory metabolomic factors and findings from recent studies that provide context and explanation for these associations in our descriptive study.

Rs17300539 is located in the promoter region of the ADIPOZ gene, which encodes adiponectin [89]. The high-risk allele has been associated with insulin resistance, whereas the low-risk allele may be associated with protection from weight regain postweight loss intervention [90]. Moreover, the high-risk allele has been associated with higher weight, BMI, and waist and hip circumferences. However, genotype-related differences in BMI became undetectable in the interaction with a diet that is low, below the median (ie, <13% of energy intake) in monounsaturated fats (MUFAs) [91]. This led researchers to propose the possibility of moderating high risk with dietary interventions to reduce MUFAs for those with the risk alleles. Of the 392 participants, 334 were homozygous for the high-risk allele (G), 54 were heterozygous for the risk allele, and 4 were homozygous for the low-risk allele (Table 3).

Table 3. Risk allele distribution of highlighted single-nucleotide polymorphisms from weight loss models.[a]

SNP[b] id	Number of participants, n (%)				
	Risk allele	0 risk allele	1 risk allele	2 risk allele	Missing
rs17300539	G	4 (1.0)	54 (13.8)	334 (85.2)	0 (0.0)
rs2016520	C	244 (62.2)	127 (32.4)	20 (5.1)	1 (0.3)
rs4074995	A	159 (63.3)	67 (26.7)	25 (10.0)	0 (0.0)

[a]Distribution of risk alleles of single-nucleotide polymorphisms from weight loss models that were highlighted in plots and *Discussion* section of the paper.

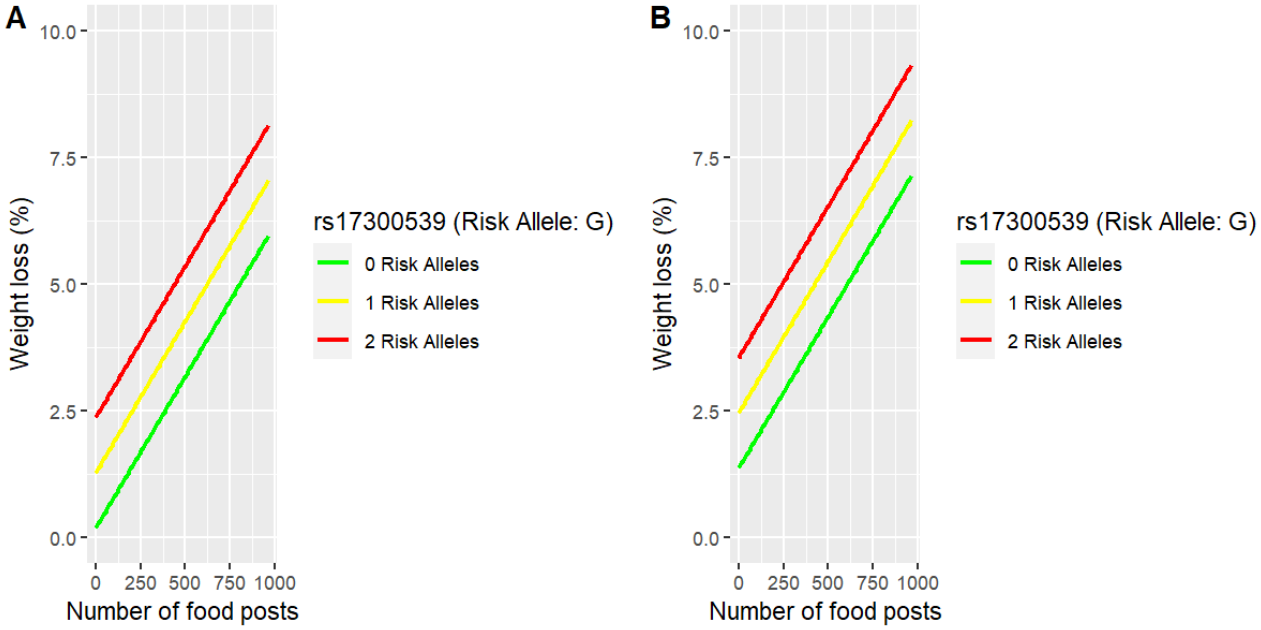
[b]SNP: single-nucleotide polymorphism.

The regression models are interpretable models describing weight loss in this cohort and may be visualized to gain insight into variables found to be significant. Figure S3 in Multimedia Appendix 1 and Figure 2 depict the relationships of engagement variables and rs17300539 to weight loss in the genomic-enhanced weight loss percent model. These plots reveal the least squares fit of weight loss percent for females (panel A) and males (panel B), as the two visualized predictors are varied while holding all other model variables constant (SNPs were held constant at their most frequent [mode] values, whereas engagement variables were held constant at their gender-specific means, except for coaching sessions completed, which was fixed at its gender-specific median). The visualizations permit us to see the model relationships of particular predictors, as they impact the outcome variable. For example, the weight loss (%) model fit to this cohort describes the average male having 2 risk alleles who posts no food photos as losing 3.5% of body weight, whereas the average male with the same genomic risk who posts 975 food photos loses 8.75% of body weight (Figure 2). The coefficients of the fitted predictors reveal that in this model, for every 100 additional food posts, participants lose, on average, an additional 0.60% weight while holding all other model predictors constant (Table S9 in Multimedia Appendix 1).

In this model, for each additional risk allele (G) of rs17300539, participants lose, on average, an additional 1.09% weight while holding all other model predictors constant (Table S9 in Multimedia Appendix 1). Similarly, as the number of risk alleles of rs17300539 increases from 0 to 1 to 2, so does percentage weight loss as a function of greater behavioral engagement measured both in the number of completed coaching sessions (Figure S3 in Multimedia Appendix 1) and the number of food photos posted (Figure 2). In essence, participants in this cohort who were at higher risk lost a greater percentage of weight compared with their lower risk counterparts. Moreover, the percentage of weight loss increased in proportion to greater behavioral engagement.

XSL•FO
RenderX

Figure 2. Weight loss (%) versus food posts by rs17300539 (monounsaturated fat intake and weight gain tendency single-nucleotide polymorphism) in females (A) compared with males (B).



Rs2016520 is a variant of the PPARD gene, which encodes a protein implicated in fat metabolism and baseline cholesterol levels [92]. In women, this SNP has been shown to be associated with muscle development and blood cholesterol reduction after a 12-week exercise regime [93]. High-risk alleles predisposed women to less weight loss on exercise [93]. Of the 392 participants, 20 were homozygous for the high-risk allele (C) of SNP rs2016520, 127 were heterozygous for the risk allele, 244 were homozygous for the low-risk allele, and 1 had no available data (Table 3).

Similar to Figure 2 and Figure S3 in Multimedia Appendix 1, Figure 3 and Figure 4 and Figure S4 in Multimedia Appendix 1 reveal the least squares fit of weight loss in pounds for females (panel A) and males (panel B) as the two visualized predictors are varied while holding all other model variables constant. The weight loss pounds model fit to this cohort describes the average female having 2 risk alleles who posts 975 food photos as losing 21 lb (9.5 kg), but only 8 lb (3.6 kg) if no food photos are posted (Figure S4A in Multimedia Appendix 1). Similarly, this model describes the average male having 2 risk alleles as losing 25 lb (11.3 kg) with 975 food posts but only 13 lb (5.9 kg) with no food posts (Figure S4B in Multimedia Appendix 1). As visualized in Figures 3 and 4 and Figure S4 in Multimedia Appendix 1 the effect in this descriptive model of an increase in the number of risk alleles from 0 to 2 is that pounds of weight loss with respect to engagement increases when engagement is measured either as the number of coaching sessions or as the number of food photos posted in the Digbi Health app. Moreover, as seen in Figure 4, a higher baseline BMI was associated with more pounds lost. Males lost more weight than females in each risk group of this SNP.

The rs4074995 SNP has been implicated in calcium-potassium regulation [94]; it is located within the RGS14 gene and is associated with both serum phosphate [95] and serum calcium [96] levels. In particular, each copy of the A allele is correlated with an increase in serum calcium concentration [96]. For the rs4074995_A SNP, of the 251 participants, 25 were homozygous for the high-risk allele (A), 67 were heterozygous for the risk allele, and 159 were homozygous for the low-risk allele (Table 3). The sample size of 251 was smaller than for the abovementioned linear models because rs4074995 was chosen to be highlighted as a predictor of the genomic-enhanced logistic regression (success vs failure) model, which was fit to a subset of the cohort that experienced success, defined as ≥5% weight loss, or failure, defined as weight gain or negligible change of <2 lb (0.9 kg).

XSL•FO
RenderX

Figure 3. Weight loss (lb) versus completed coaching sessions by rs2016520 (cholesterol single-nucleotide polymorphism) in females (A) compared with males (B).

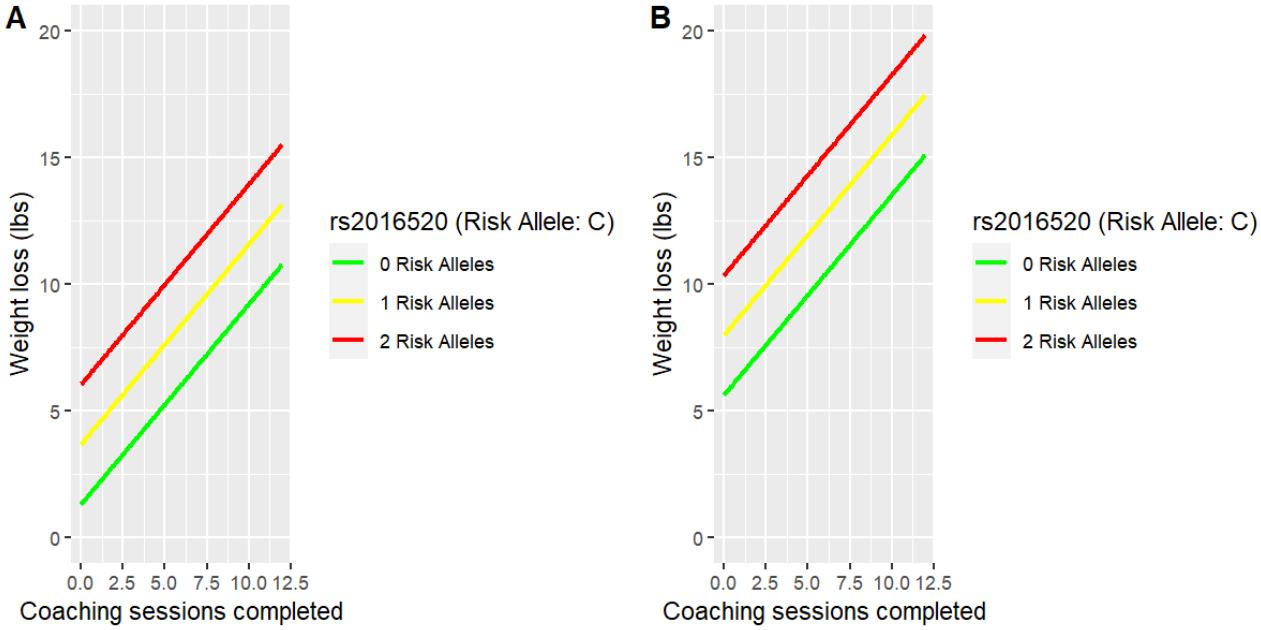
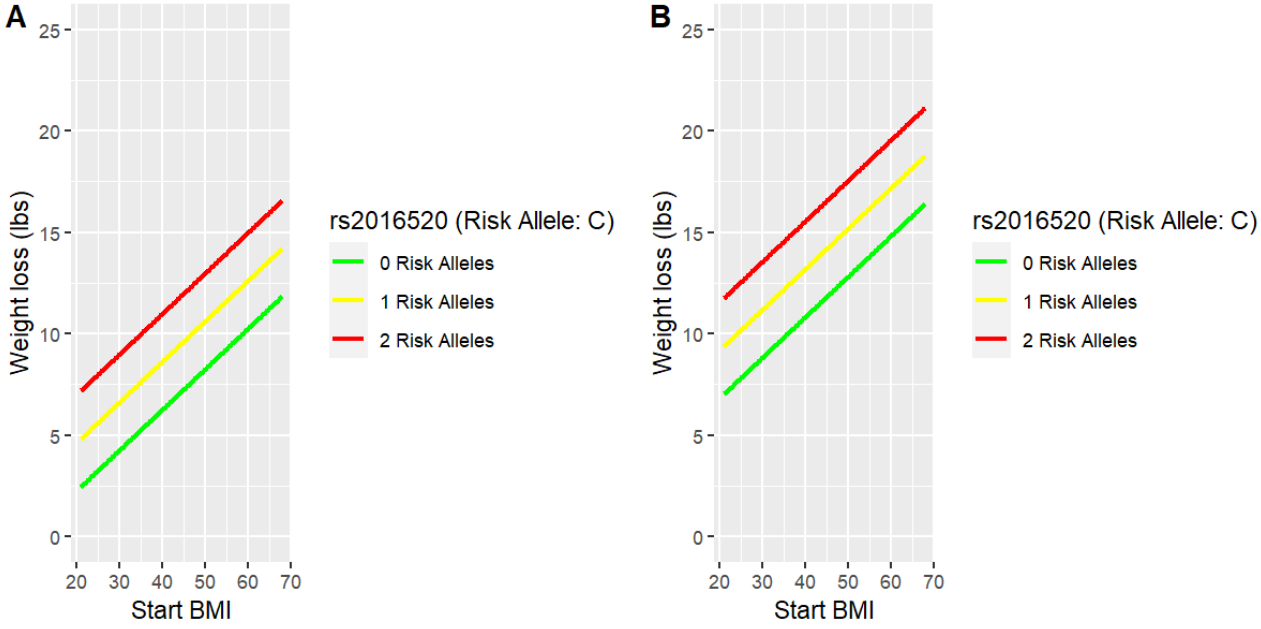


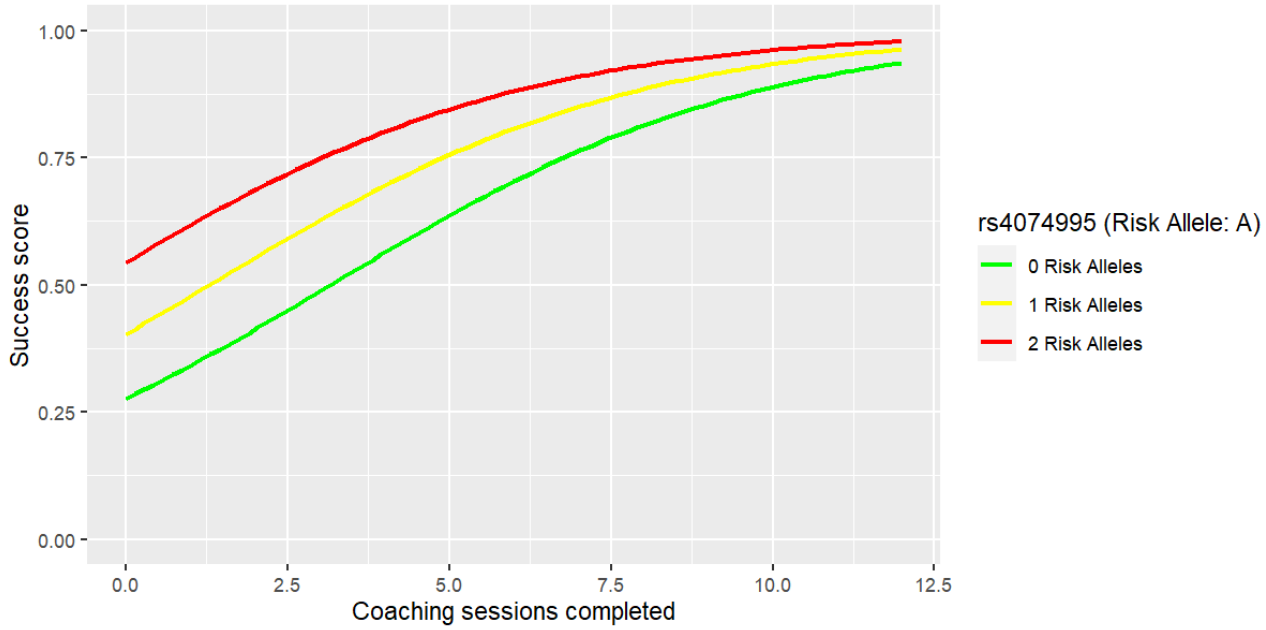
Figure 4. Weight loss (lb) versus baseline BMI by rs2016520 (cholesterol SNP) in females (A) compared with males (B).



As seen in Figure 5 and Figure S5 in Multimedia Appendix 1, as the risk alleles of SNP rs4074995 increase from 0 to 1 to 2, there is an increase in the success score, which is the likelihood of this model assigning a particular observation to the success class. Similar to the other models, we found that an increased number of coaching sessions completed is associated with a sharp increase in the success score. However, on highest engagement, the effect of risk status diminishes (those with 0, 1, or 2 risk alleles were approximately equally likely to achieve weight loss on highest engagement).

XSL•FO
RenderX

Figure 5. Weight loss success versus completed coaching sessions by rs4074995 (calcium-potassium single-nucleotide polymorphism).



Discussion

Principal Findings

A group of 393 participants underwent lifestyle changes over 120 days through the Digbi Health program—a precision digital care program applying machine learning analytics to genetic and microbiome profiles, demographics, and self-reported lifestyle habits—delivering care through the app and weekly health coaching check-ins. Over the duration of the program, patients' genomic and gut microbiome data pertinent to weight loss (from Digbi Health–curated panels) were provided and translated into lifestyle recommendations and recipes. Of the participants, 72% (283/393) lost weight, whereas 10.7% (42/393) gained ≥2 lb. Of those who lost weight, 50.2% (142/283) were able to lose 5% or more over 120 days.

Interpretable linear regression models of weight loss in this cohort (pounds lost and percentage lost) as a function of demographic and behavioral engagement variables were fit to describe the weight loss of this cohort. Genomic-enhanced models were also built by adding participant genomic data as predictors. Interpretable success or failure logistic regression models were also fit, with and without genomic data. The addition of genomic predictors substantially improved the fit of all models.

The fitted models were examined to gain insights into the weight loss journey of this cohort. Gender, engagement, and specific SNP risk alleles were significantly associated with successful weight loss. The models described greater average weight loss in our cohort for participants having more of certain risk alleles. Here, we consider how successful weight loss may be obtained in the face of greater genetic risk factors. Notably, Digbi Health precision coaching for lifestyle modification is personalized to these genetic risks, and patients reported realizing success that was previously unattainable after being empowered by the knowledge of their genetic and microbiome risk factors, accompanied by advice on lifestyle modifications to address these risks.

We profiled three of these genetic markers (see Results section) to elucidate their relationships with explanatory metabolomic processes and weight gain and loss. Here, we connected these relationships with personalized recommendations delivered by the Digbi Health app and coaching staff. The profiled SNPs were associated with circulating adiponectin and response to dietary MUFA consumption, fat metabolism, and baseline cholesterol levels, and serum calcium levels and calcium-potassium metabolism were strongly associated with weight loss success.

As an example of personalized dietary advice delivered by both the app and coach for program participants who are at genetic risk of weight gain, we considered the advice delivered to participants with a different genetic outlook with regard to rs17300539, a risk allele for weight gain with high MUFA intake. This SNP is depicted in the visualizations of our linear model for weight loss percentage (Figure 2 and Figure S3 in Multimedia Appendix 1). As reported earlier, participants in this cohort who were at higher risk lost a greater percentage of weight compared with their lower risk counterparts, and percentages of weight loss correlated with greater behavioral engagement. This finding can be explained by the fact that those with high risk for this trait were advised by both the app and human coaching to avoid MUFA consumption as much as possible (contrary to the conventional wisdom that these—olive oils, almond oils, etc—are comparatively healthy fats). Instead, they were advised to shift to the consumption of polyunsaturated or saturated fats, depending on their genotypes [97]. Moreover, this SNP is associated with insulin resistance, and parts of the Digbi Health Nutrition Plan (eg, intermittent fasting and reducing processed carbohydrate consumption) would be expected to reduce insulin resistance, addressing the risk associated with this SNP, thereby helping with weight loss [93].

XSL•FO
RenderX

The linear regression pounds lost models found an association between higher baseline BMI and increased weight loss in this cohort (Figure 4). For each one-unit increase in baseline BMI, participants lost an additional 0.2 lb (0.09 kg) on average while holding the other variables in the model constant. This finding could be encouraging to new participants with higher BMI, who may have attempted weight loss with other programs but without much success. Adding genomic information (Figure 3), the model describes the average man of this cohort at the highest baseline BMI, having 2 risk alleles for rs2016520_C and completing five coaching sessions as losing 21 lb (9.5 kg), but women having the same risk outlook and behavioral engagement as losing, on average, 16.5 lb (7.5 kg). When compared with participants of the same gender, baseline BMI, number of coaching sessions, and genomic outlook for all SNPs except rs2016520_C, participants in this data set lost 2.4 lb (1.1 kg), on average, over their treatment for each additional risk allele they had of rs2016520_C (Table S7 in Multimedia Appendix 1).

This SNP is poorly characterized in the general population, but studies associate it with BMI and waist circumference among Han Chinese [98] as well as with cholesterol metabolism [99]. This latter association drives the recommendation by Digbi Health that participants presenting with the high-risk allele limit cholesterol consumption. The association between the number of risk alleles of rs2016520_C and increased weight loss in this data set may indicate the efficacy of data-driven coaching by Digbi Health. A visualization in Figure 3 of this descriptive model depicts a male of average baseline BMI and the most frequent genomic outlook for all except rs2016520_C as losing 20 lb (9.0 kg), on average, if he completed 12 coaching sessions, but only 11 lb (5.0 kg) on average with only one coaching session over the course of treatment.

Calcium is an essential mineral critical for vascular function, muscle function, neurotransmission, cell signaling, and hormone secretion [100]. Serum calcium levels tend not to respond directly to dietary calcium intake, and instead, the body relies on reservoirs in bone tissue to maintain consistent calcium concentrations [100]. Recent research has emerged tying higher serum calcium levels to the development of insulin resistance and cardiovascular hypertension [101]. High serum calcium levels have long been correlated with obesity [102]. Figure 5 depicts the success or fail logistic regression model of the associations between the number of coaching sessions completed and rs4074995_A with successful weight loss while holding all other variables in the model constant at their most frequent number of risk alleles. As with the linear model, in this Digbi Health treatment cohort, increasing total coaching sessions was associated with higher success in losing weight. Those at high risk for excess serum calcium levels were especially encouraged to embrace intermittent fasting and carbohydrate avoidance to combat insulin resistance. This may explain their higher success in achieving ≥5% weight loss (Figure S5 in Multimedia Appendix 1). We found that for participants with more risk alleles of the rs4074995 SNP, success in weight loss increased with more coaching, although it was not as pronounced in those with minimum (0) risk alleles. It may be that success for those with more risk alleles was not as heavily dependent on more coaching sessions, as the app itself conveys pertinent dietary advice.

In addition, our data strongly indicate that behavioral engagement, particularly coaching, contributed to weight loss success. Participants experienced, on average, 0.37% more weight loss with each additional coaching session while holding all other model variables constant (Table S9 in Multimedia Appendix 1). All models found weight loss to be significantly associated with behavioral engagement with the program and app (number of coaching sessions completed, weight entries logged, and food photos logged as predictors). Food photos and weight tracking showed more than 98% correlation with each other and both were significantly associated with weight loss success (Tables S6 to S11 in Multimedia Appendix 1). Prior research has shown that regular engagement with digital weight loss platforms and regular weight tracking is associated with greater weight loss success [103].

We hypothesized that successful weight loss was achieved by adhering to data-driven dietary recommendations that depart from conventional nutritional weight loss advice. Of those 10.7% (42/393) of the participant population who gained weight, there was a notable lack of engagement in the program. Those who gained weight, compared with their counterparts who lost weight, tended to neither engage in coaching nor regularly use the Digbi Health app to log body weight and post food photos of meals. Those who checked in with the coach regularly and logged into the app frequently to post weight and food photos were more likely to lose weight than those who did not.

Coaching sessions completed, along with other behavioral engagement variables, differed between participants who lost weight and those who gained weight, whereas baseline weight and BMI did not. The density plots in Figure 6 fairly compare distributions of the two groups: although many more people lost weight than gained, the area under the curve of each group is uniform at 1. Figure 6A illustrates the distributions of completed coaching sessions for those who lost weight (blue) versus those who gained weight (red). The difference is striking: only a fraction of those who failed to lose weight completed at least five (the mean and median) coaching sessions, whereas those who succeeded generally completed five or more. All three measures of engagement were significantly higher in participants who lost weight (blue) versus those who gained weight (red). These distributions are visualized in Figure 6. In contrast, however, Figure S6 in Multimedia Appendix 1 shows no statistical difference in means in (A) baseline weight and (B) baseline BMI, confirmed by the Welch two-sided two-sample *t* test (P=.64 and P=.42, respectively) between participants who lost (blue) versus gained (red) weight.

XSL•FO
RenderX

Figure 6. Distributions of engagement variables differ by weight loss group. Measures of engagement were higher in participants who lost weight (blue) versus those who gained weight (red). Statistical difference in means confirmed by the Welch two-sample t test (A) coaching sessions (P<.001), (B) number of weight entries (P<.001), and (C) number of food posts (P<.001). Less than 2-lb gain or loss was considered negligible and excluded from this figure. Engagement variables were summed over the study period of 120 days.

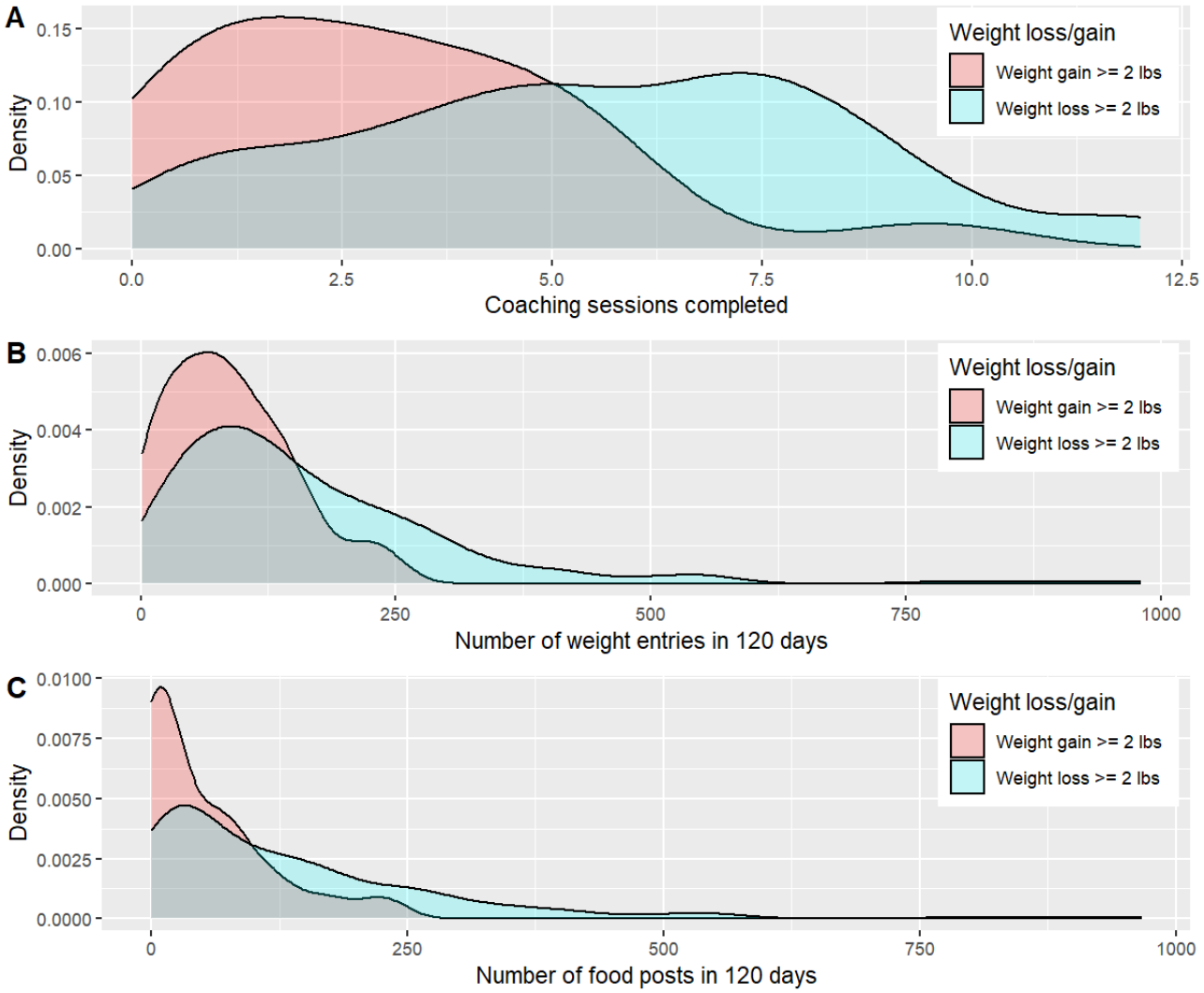


Table 2 and Table S1 in Multimedia Appendix 1 show the distributions of variables based on gender. A notable feature of this cohort is that women are grossly overrepresented—a feature that is not specific to the demographics of obesity. Although globally, more women are obese than men, the disparity is driven in large part by demographics, particularly in Africa and the Middle East. In Western countries, men are more likely to be obese [104], which is not reflected in our sample. Instead, our participant demographics may be more reflective of individual self-image. Women appear more likely to perceive themselves as overweight and are more likely to attempt weight loss [105]. Exploration of gender differences in weight loss maintenance reported that men comprised only 27% of participants in behavioral weight loss programs [106,107].

Additionally pertinent to the gender composition of this cohort is that patients self-select and continue with the weight management program of Digbi Health based on the approach they feel works for them. One study indicated that women were more than twice as likely to report having used an organized weight loss program, whereas men were more self-directed in their weight loss [108]. In a single-blinded, randomized clinical trial, the efficacies of three e-coaching approaches were compared: no coaching, nondirected coaching, and directed coaching. Women achieved most success (weight loss, reduction of waist circumference, and improvement of physical activity) in the first 12 weeks with directive e-coaching (similar to Digbi Health), whereas men lost more weight with nondirected e-coaching [80,109]. Men who chose the Digbi Health program succeeded in weight loss. In this cohort, men lost more weight than women over 120 days, consistent with previous findings that men lose weight faster than women [110]. However, the low participation in behavioral weight loss programs, such as that examined here, points to the importance of identifying the underlying factors that impact early engagement and success in weight loss. We are currently undertaking such a study, with the aim of innovating for a level of personalization that will empower and drive success with different subgroups of clients, including men.

Recent research has explored the incorporation of genotypic information into nutritional advice. Researchers have tested the hypothesis that dietary interventions using personalized genotype information have greater efficacy than the same

XSL•FO
RenderX

interventions without genomic data, achieving mixed findings [111-115]. However, these studies tested only a few SNPs. Some groups found no benefit from genomic information [22,111], whereas others found specific SNPs (eg, APOE and ACE genes) to be associated with improved dietary changes [21,23]. Some studies observed an increase in positive dietary behaviors on the part of participants who carried a risk allele pertinent to a dietary factor (eg, sodium and fat) with whose recommendations they had not been adhering [21,23]. This aligns with our anecdotal experience of Digbi Health clients, including those in this cohort who had risk alleles for rs17300539, rs2016520, and rs4074995, as described earlier.

Food4Me [111] conducted an RCT to investigate the effectiveness of internet-based personalized nutrition interventions on weight loss and dietary intake with 1269 participants randomized into four groups: one control without personalization and three personalized groups, each with an additional level of personalization: personalized by individual baseline diet, by baseline diet and phenotype (anthropometry and blood markers), and by baseline diet, phenotype, and genotype (5 SNPs). The group comprising participants combined from all three levels of personalized nutrition experienced significantly better improvements in body weight and BMI at month 3 and a more positive behavioral change than the control group. The authors found no evidence that the addition of either phenotype or phenotype plus genotype information to individual baseline diet data enhanced the efficacy of the personalized intervention. However, none of the three personalization interventions, including individual baseline diet, were reportedly tested in isolation for differences in outcomes from the control protocol, and furthermore only five variants were used in genomic personalization.

The five variants used in the Food4Me genomic personalization were from the MTHFR, FTO, TCF7L2, APOEε4, and FADS1 genes [116]. Of these five variants, three were present in the Digbi Health panel of 197 curated SNPs, and of these 3 SNPs, only rs7903146 from the TCF7L2 gene was found to be significant to any of our 3 genomic-enhanced models (linear percent weight loss, linear pounds lost, and logistic success or fail). Rs9939609 (FTO gene) and rs1801133 (MTHFR) were the 2 SNPs that were a part of our gene panel, and hence included as variables, but were found to not significantly contribute to any of the three weight loss models. Rs7903146 was found to be significant to the linear percent weight loss model, and although it was selected as a variable by Lasso for the pounds lost model, it was not statistically significant. Research associates rs7903146 with a higher risk of gestational [117,118] and type 2 [119-121] diabetes as well as reduced insulin levels [119,122]. These medical implications are especially relevant to postmenopausal women [123] as well as those in childbearing years. Unlike the pounds lost model, the percent weight loss model captures weight loss independently of start weight and more aptly models women's weight loss alongside that of men. Digbi Health has a robust protocol to address diabetes risk and insulin resistance, which differs substantially from the Food4Me personalization based on rs7903146 as reflected in the Food4Me article (including the decision tree for TCF7L2-based information delivered to level 3 "*Diet plus phenotype plus genotype*" in Figure S3 in Multimedia Appendix 1) [111]; thus, we would not expect the outcomes of our personalization regarding this SNP to be the same. We noted that Digbi Health personalization arises from the much broader set of traits in our curated panels and a greater number of SNPs associated with those traits, enabling personalization that is more fine-grained than that informed by a few SNPs. We are not aware of any RCT to test the added value of genomic data that uses the breadth of genotypic markers considered in the Digbi Health program.

Conclusions

Over the last two decades, the obesity epidemic has coincided with a dramatic change in unhealthy eating habits, a sedentary lifestyle, and physical inactivity. In the United States, more than 40% of the adult population is now overweight or obese. Hereditary predisposition to obesity may have interacted with the obesogenic environment and contributed even further toward the epidemic. The recent accumulation of genomic and lifestyle data has led to the demonstration of possible effects of gene-environmental interactions on obesity [124]. Data from dietary intervention trials indicate that genetic variants, particularly those linked to obesity, metabolism, and nutrient consumption, may significantly alter changes in adiposity and metabolic responses to nutritional interventions and promote effective weight loss [59].

In the foreseeable future, the incorporation of data on genes, eating patterns, metabolites, and gut microbiome into weight loss interventions will be one of the most promising fields of precision care and may allow for the generation of predictable weight loss models based on individual genomic, microbiomic, and metabolomic factors. The goal is precision nutrition, individually tailored to enable effective weight loss and prevent chronic diseases on the basis of genomic history; habitual consumption of food and drink; intake of nutrients (especially those that contribute to disease risks); and metabolomics, microbiome, and other omics profiles of a person [59].

Although using precision medicine to target heterogeneous conditions may seem counter-intuitive, it is the heterogeneous nature of conditions, such as obesity and metabolic illness, that make them such potent targets for intervention, impacting the greatest number of people [8]. Obese subpopulations identified as genetically predisposed to favorably or unfavorably respond to a given weight loss intervention could be targeted accordingly.

To date, few studies have investigated metabolomic functioning, lifestyle and behavioral mechanisms, and gut microbiome, which can affect obesity and health at the interface between genetic variation and the environment. The Digbi Health digital precision weight loss program operates at this interface. This study was limited by its retrospective and descriptive nature. The field of precision nutrition would benefit from additional prospective randomized controlled studies on a larger scale. Although such studies will be needed to validate these findings, the analysis and modeling presented here appear to support dietary precision interventions considering genetic predisposition to disease and genetic variants defining dietary preference and metabolic risk. In addition, our results point to the efficacy of

XSL•FO
RenderX

coaching that empowers and actively engages participants in their own success.

Future studies should explore the synergistic effects of genomic variables in interactions with other genome, microbiome, and lifestyle and behavior variables. A follow-up to the work presented here, exploring not only the effect of incorporating genomic data but also including the microbiome data used in Digbi Health precision care, is currently in preparation. Personalized protocols that incorporate data on genes, eating patterns, metabolites, and gut microbiome into weight loss interventions may well be a promising field of precision care, allowing for the generation of predictable weight loss models that account for the synergistic effect of these influential factors.

Acknowledgments

This study was funded by Digbi Health, Mountain View, California, United States.

Conflicts of Interest

Digbi Health is sponsoring this study, and the principal investigator and study staff have a financial interest in the company. RS is the founder and CEO of Digbi Health. Authors DK, RRR, SS-R, KMM, VS, CI, CR-S, GS, and PAF-L are employees at Digbi Health. IJ is the Chief Medical Officer employed by Digbi Health.

Multimedia Appendix 1
Distributions of predictors, weight loss outcomes, and results of descriptive models.
[PDF File (Adobe PDF File), 363 KB - jmir_v23i5e25401_app1.pdf]

References

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE). Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-10-18]
2. Cunningham JW, Vaduganathan M, Claggett BL, Jering KS, Bhatt AS, Rosenthal N, et al. Clinical outcomes in young US adults hospitalized with COVID-19. JAMA Intern Med 2020 Sep 09;181(3):379. [doi: 10.1001/jamainternmed.2020.5313] [Medline: 32902580]
3. Prevalence of obesity and severe obesity among adults: United States, 2017–2018. NCHS Data Brief No. 360, February 2020. URL: https://www.cdc.gov/nchs/products/databriefs/db360.htm [accessed 2020-10-19]
4. Townsend MJ, Kyle TK, Stanford FC. Outcomes of COVID-19: disparities in obesity and by ethnicity/race. Int J Obes (Lond) 2020 Sep 09;44(9):1807-1809 [FREE Full text] [doi: 10.1038/s41366-020-0635-2] [Medline: 32647359]
5. Apovian C. Obesity: definition, comorbidities, causes, and burden. Am J Manag Care 2016 Jun;22(7 Suppl):176-185 [FREE Full text] [Medline: 27356115]
6. McAllister EJ, Dhurandhar NV, Keith SW, Aronne LJ, Barger J, Baskin M, et al. Ten putative contributors to the obesity epidemic. Crit Rev Food Sci Nutr 2009 Nov 02;49(10):868-913 [FREE Full text] [doi: 10.1080/10408390903372599] [Medline: 19960394]
7. Apovian CM, Aronne LJ, Bessesen DH, McDonnell ME, Murad MH, Pagotto U, Endocrine Society. Pharmacological management of obesity: an endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2015 Feb;100(2):342-362. [doi: 10.1210/jc.2014-3415] [Medline: 25590212]
8. Bray MS, Loos RJ, McCaffery JM, Ling C, Franks PW, Weinstock GM, Conference Working Group. NIH working group report-using genomic information to guide weight management: from universal to precision treatment. Obesity (Silver Spring) 2016 Jan 25;24(1):14-22 [FREE Full text] [doi: 10.1002/oby.21381] [Medline: 26692578]
9. Joo JK, Lee KS. Pharmacotherapy for obesity. J Menopausal Med 2014 Dec;20(3):90-96 [FREE Full text] [doi: 10.6118/jmm.2014.20.3.90] [Medline: 25580419]
10. Severin R, Sabbahi A, Mahmoud AM, Arena R, Phillips SA. Precision medicine in weight loss and healthy living. Prog Cardiovasc Dis 2019 Jan;62(1):15-20 [FREE Full text] [doi: 10.1016/j.pcad.2018.12.012] [Medline: 30610881]
11. Saad L. To lose weight, Americans rely more on dieting than exercise. Gallup Poll Social Series: Health and Healthcare. 2011. URL: https://news.gallup.com/poll/150986/Lose-Weight-Americans-Rely-Dieting-Exercise.aspx [accessed 2020-10-19]
12. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Writing Group of the PREMIER Collaborative Research Group. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. J Am Med Assoc 2003 Apr 23;289(16):2083-2093. [doi: 10.1001/jama.289.16.2083] [Medline: 12709466]
13. Diabetes Prevention Program Research Group. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 2009 Nov;374(9702):1677-1686. [doi: 10.1016/s0140-6736(09)61457-4]
14. Parkman K. Jenny Craig vs. nutrisystem vs. weight watchers. Consumer Affairs. URL: https://www.consumeraffairs.com/health/jenny-craig-vs-nutrisystem-vs-weight-watchers.html [accessed 2020-10-19]
15. Brennan L. Metabolomics in nutrition research–a powerful window into nutritional metabolism. Essays Biochem 2016;60(5):451-458. [doi: 10.1042/ebc20160029]

XSL•FO
RenderX

16. Allam-Ndoul B, Guénard F, Garneau V, Cormier H, Barbier O, Pérusse L, et al. Association between metabolite profiles, metabolic syndrome and obesity status. Nutrients 2016 May 27;8(6):324 [FREE Full text] [doi: 10.3390/nu8060324] [Medline: 27240400]
17. Bakker G, van Erk MJ, Pellis L, Wopereis S, Rubingh C, Cnubben N, et al. An antiinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: a nutrigenomics approach. Am J Clin Nutr 2010 Apr;91(4):1044-1059. [doi: 10.3945/ajcn.2009.28822] [Medline: 20181810]
18. Paquette M, Medina Larqué AS, Weisnagel SJ, Desjardins Y, Marois J, Pilon G, et al. Strawberry and cranberry polyphenols improve insulin sensitivity in insulin-resistant, non-diabetic adults: a parallel, double-blind, controlled and randomised clinical trial. Br J Nutr 2017 Mar 14;117(4):519-531. [doi: 10.1017/s0007114517000393]
19. Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, Olsson AH, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet 2013 Jun 27;9(6):e1003572 [FREE Full text] [doi: 10.1371/journal.pgen.1003572] [Medline: 23825961]
20. Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, et al. Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. Diabetes 2012 Dec 01;61(12):3322-3332 [FREE Full text] [doi: 10.2337/db11-1653] [Medline: 23028138]
21. Jacobsen SC, Gillberg L, Bork-Jensen J, Ribel-Madsen R, Lara E, Calvanese V, et al. Young men with low birthweight exhibit decreased plasticity of genome-wide muscle DNA methylation by high-fat overfeeding. Diabetologia 2014 Jun 26;57(6):1154-1158. [doi: 10.1007/s00125-014-3198-8] [Medline: 24570141]
22. Rodgers GP, Collins FS. Precision nutrition-the answer to "What to Eat to Stay Healthy". J Am Med Assoc 2020 Aug 25;324(8):735-736. [doi: 10.1001/jama.2020.13601] [Medline: 32766768]
23. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, MAGIC, Procardis Consortium, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010 Nov 10;42(11):937-948 [FREE Full text] [doi: 10.1038/ng.686] [Medline: 20935630]
24. Fall T, Ingelsson E. Genome-wide association studies of obesity and metabolic syndrome. Mol Cell Endocrinol 2014 Jan 25;382(1):740-757. [doi: 10.1016/j.mce.2012.08.018] [Medline: 22963884]
25. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, LifeLines Cohort Study, ADIPOGen Consortium, AGEN-BMI Working Group, CARDIOGRAMplusC4D Consortium, CKDGen Consortium, GLGC, ICBP, MAGIC Investigators, MuTHER Consortium, MIGen Consortium, PAGE Consortium, ReproGen Consortium, GENIE Consortium, International Endogene Consortium, et al. . [doi: 10.1038/nature14177] [Medline: 25673413]
26. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, DIAbetes Genetics ReplicationMeta-analysis (DIAGRAM) Consortium, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012 Sep 12;44(9):991-1005 [FREE Full text] [doi: 10.1038/ng.2385] [Medline: 22885924]
27. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature 2016 Aug 04;536(7614):41-47 [FREE Full text] [doi: 10.1038/nature18642] [Medline: 27398621]
28. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet 2005 Apr;6(4):287-298. [doi: 10.1038/nrg1578] [Medline: 15803198]
29. Fabbrini E, Yoshino J, Yoshino M, Magkos F, Luecking CT, Samovski D, et al. Metabolically normal obese people are protected from adverse effects following weight gain. J Clin Invest 2015 Jan 2;125(2):787-795. [doi: 10.1172/jci78425]
30. Heianza Y, Arase Y, Tsuji H, Fujihara K, Saito K, Hsieh SD, et al. Metabolically healthy obesity, presence or absence of fatty liver, and risk of type 2 diabetes in Japanese individuals: Toranomon Hospital Health Management Center Study 20 (TOPICS 20). J Clin Endocrinol Metab 2014 Aug;99(8):2952-2960. [doi: 10.1210/jc.2013-4427] [Medline: 24823457]
31. Roberson LL, Aneni EC, Maziak W, Agatston A, Feldman T, Rouseff M, et al. Beyond BMI: The "Metabolically healthy obese" phenotype & its association with clinical/subclinical cardiovascular disease and all-cause mortality -- a systematic review. BMC Public Health 2014 Jan 08;14:14 [FREE Full text] [doi: 10.1186/1471-2458-14-14] [Medline: 24400816]
32. Zheng R, Zhou D, Zhu Y. The long-term prognosis of cardiovascular disease and all-cause mortality for metabolically healthy obesity: a systematic review and meta-analysis. J Epidemiol Community Health 2016 Oct 28;70(10):1024-1031. [doi: 10.1136/jech-2015-206948] [Medline: 27126492]
33. Bell JA, Kivimaki M, Hamer M. Metabolically healthy obesity and risk of incident type 2 diabetes: a meta-analysis of prospective cohort studies. Obes Rev 2014 Jun 24;15(6):504-515 [FREE Full text] [doi: 10.1111/obr.12157] [Medline: 24661566]
34. Courcoulas AP, Yanovski SZ, Bonds D, Eggerman TL, Horlick M, Staten MA, et al. Long-term outcomes of bariatric surgery: a National Institutes of Health symposium. JAMA Surg 2014 Dec 01;149(12):1323-1329 [FREE Full text] [doi: 10.1001/jamasurg.2014.2440] [Medline: 25271405]
35. Madsbad S, Dirksen C, Holst JJ. Mechanisms of changes in glucose metabolism and bodyweight after bariatric surgery. Lancet Diabetes Endocrinol 2014 Feb;2(2):152-164. [doi: 10.1016/s2213-8587(13)70218-3]
36. Carlsson LM, Sjöholm K, Karlsson C, Jacobson P, Andersson-Assarsson JC, Svensson P, et al. Long-term incidence of microvascular disease after bariatric surgery or usual care in patients with obesity, stratified by baseline glycaemic status:

XSL•FO
RenderX

a post-hoc analysis of participants from the Swedish Obese Subjects study. Lancet Diabetes Endocrinol 2017 Apr;5(4):271-279. [doi: 10.1016/s2213-8587(17)30061-x]
37. Hatoum IJ, Greenawalt DM, Cotsapas C, Reitman ML, Daly MJ, Kaplan LM. Heritability of the weight loss response to gastric bypass surgery. J Clin Endocrinol Metab 2011 Oct;96(10):1630-1633 [FREE Full text] [doi: 10.1210/jc.2011-1130] [Medline: 21832118]
38. Rinella E, Still C, Shao Y, Wood GC, Chu X, Salerno B, et al. Genome-wide association of single-nucleotide polymorphisms with weight loss outcomes after Roux-en-Y gastric bypass surgery. J Clin Endocrinol Metab 2013 Jun;98(6):1131-1136 [FREE Full text] [doi: 10.1210/jc.2012-3421] [Medline: 23633212]
39. Hatoum I, Greenawalt D, Cotsapas C, Daly M, Reitman M, Kaplan L. Weight loss after gastric bypass is associated with a variant at 15q26.1. Am J Hum Genet 2013 May 02;92(5):827-834 [FREE Full text] [doi: 10.1016/j.ajhg.2013.04.009] [Medline: 23643386]
40. Breen FM, Plomin R, Wardle J. Heritability of food preferences in young children. Physiol Behav 2006 Jul 30;88(4-5):443-447. [doi: 10.1016/j.physbeh.2006.04.016] [Medline: 16750228]
41. Törnwall O, Silventoinen K, Hiekkalinna T, Perola M, Tuorila H, Kaprio J. Identifying flavor preference subgroups. Genetic basis and related eating behavior traits. Appetite 2014 Apr;75:1-10. [doi: 10.1016/j.appet.2013.11.020] [Medline: 24361469]
42. Wardle J, Carnell S, Haworth C, Farooqi I, O'Rahilly S, Plomin R. Obesity associated genetic variation in FTO is associated with diminished satiety. J Clin Endocrinol Metab 2008 Sep;93(9):3640-3643. [doi: 10.1210/jc.2008-0472] [Medline: 18583465]
43. Timpson N, Emmett P, Frayling T, Rogers I, Hattersley AT, McCarthy MI, et al. The fat mass- and obesity-associated locus and dietary intake in children. Am J Clin Nutr 2008 Oct;88(4):971-978 [FREE Full text] [doi: 10.1093/ajcn/88.4.971] [Medline: 18842783]
44. McCaffery J, Papandonatos G, Peter I, Huggins GS, Raynor HA, Delahanty LM, Genetic Subgroup of Look AHEAD, Look AHEAD Research Group. Obesity susceptibility loci and dietary intake in the Look AHEAD Trial. Am J Clin Nutr 2012 Jun;95(6):1477-1486 [FREE Full text] [doi: 10.3945/ajcn.111.026955] [Medline: 22513296]
45. Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CN. An obesity-associated gene variant and increased energy intake in children. N Engl J Med 2008 Dec 11;359(24):2558-2566. [doi: 10.1056/nejmoa0803839]
46. Feeney E, O'Brien S, Scannell A, Markey A, Gibney ER. Genetic variation in taste perception: does it have a role in healthy eating? Proc Nutr Soc 2010 Nov 22;70(1):135-143. [doi: 10.1017/s0029665110003976]
47. Qi L. Gene-diet interaction and weight loss. Curr Opin Lipidol 2014 Feb;25(1):27-34 [FREE Full text] [doi: 10.1097/MOL.0000000000000037] [Medline: 24345984]
48. Qi L. Gene-diet interactions in complex disease: current findings and relevance for public health. Curr Nutr Rep 2012 Dec 01;1(4):222-227 [FREE Full text] [doi: 10.1007/s13668-012-0029-8] [Medline: 23139897]
49. Qi L, Cho Y. Gene-environment interaction and obesity. Nutr Rev 2008 Dec;66(12):684-694 [FREE Full text] [doi: 10.1111/j.1753-4887.2008.00128.x] [Medline: 19019037]
50. UK biobank. URL: http://www.ukbiobank.ac.uk/2016/11/scanning-study-launched/ [accessed 2018-09-05] [WebCite Cache ID 72CgBSM4O]
51. Lv J, Yu C, Guo Y, Bian Z, Yang L, Chen Y, China Kadoorie Biobank Collaborative Group. Adherence to healthy lifestyle and cardiovascular diseases in the Chinese population. J Am Coll Cardiol 2017 Mar 07;69(9):1116-1125 [FREE Full text] [doi: 10.1016/j.jacc.2016.11.076] [Medline: 28254173]
52. Di Pietro L, Dziura J, Blair SN. Estimated change in physical activity level (PAL) and prediction of 5-year weight change in men: the Aerobics Center Longitudinal Study. Int J Obes Relat Metab Disord 2004 Dec 29;28(12):1541-1547. [doi: 10.1038/sj.ijo.0802821] [Medline: 15543159]
53. Haapanen N, Miilunpalo S, Pasanen M, Oja P, Vuori I. Association between leisure time physical activity and 10-year body mass change among working-aged men and women. Int J Obes Relat Metab Disord 1997 Apr 18;21(4):288-296. [doi: 10.1038/sj.ijo.0800403] [Medline: 9130026]
54. Schmitz K, Jacobs D, Leon A, Schreiner P, Sternfeld B. Physical activity and body weight: associations over ten years in the CARDIA study. Coronary artery risk development in young adults. Int J Obes Relat Metab Disord 2000 Nov 17;24(11):1475-1487. [doi: 10.1038/sj.ijo.0801415] [Medline: 11126345]
55. Washburn RA, Szabo AN, Lambourne K, Willis EA, Ptomey LT, Honas JJ, et al. Does the method of weight loss effect long-term changes in weight, body composition or chronic disease risk factors in overweight or obese adults? A systematic review. PLoS One 2014 Oct 15;9(10):e109849 [FREE Full text] [doi: 10.1371/journal.pone.0109849] [Medline: 25333384]
56. Herring MP, Sailors MH, Bray MS. Genetic factors in exercise adoption, adherence and obesity. Obes Rev 2014 Jan 15;15(1):29-39 [FREE Full text] [doi: 10.1111/obr.12089] [Medline: 24034448]
57. Maia JA, Thomis M, Beunen G. Genetic factors in physical activity levels. Am J Prev Med 2002 Aug;23(2):87-91. [doi: 10.1016/s0749-3797(02)00478-6]
58. Cai G, Cole SA, Butte N, Bacino C, Diego V, Tan K, et al. A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. Obesity (Silver Spring) 2006 Sep;14(9):1596-1604 [FREE Full text] [doi: 10.1038/oby.2006.184] [Medline: 17030971]

XSL•FO
RenderX

59. de Toro-Martín J, Arsenault BJ, Després JP, Vohl M. Precision nutrition: a review of personalized nutritional approaches for the prevention and management of metabolic syndrome. Nutrients 2017 Aug 22;9(8):913 [FREE Full text] [doi: 10.3390/nu9080913] [Medline: 28829397]
60. Reinehr T, Roth CL. The gut sensor as regulator of body weight. Endocrine 2015 May 30;49(1):35-50. [doi: 10.1007/s12020-014-0518-1] [Medline: 25548085]
61. Frost F, Storck LJ, Kacprowski T, Gärtner S, Rühlemann M, Bang C, et al. A structured weight loss program increases gut microbiota phylogenetic diversity and reduces levels of Collinsella in obese type 2 diabetics: a pilot study. PLoS One 2019 Jul 18;14(7):e0219489 [FREE Full text] [doi: 10.1371/journal.pone.0219489] [Medline: 31318902]
62. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 2013 Sep 06;341(6150):1241214 [FREE Full text] [doi: 10.1126/science.1241214] [Medline: 24009397]
63. Kang JX. Gut microbiota and personalized nutrition. J Nutrigenet Nutrigenomics 2013;6(2):I-II [FREE Full text] [doi: 10.1159/000353144] [Medline: 23774142]
64. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, et al. The effect of host genetics on the gut microbiome. Nat Genet 2016 Nov 3;48(11):1407-1412. [doi: 10.1038/ng.3663] [Medline: 27694959]
65. Corella D, Arregui M, Coltell O, Portolés O, Guillem-Sáiz P, Carrasco P, et al. Association of the LCT-13910C>T polymorphism with obesity and its modulation by dairy products in a Mediterranean population. Obesity (Silver Spring) 2011 Aug;19(8):1707-1714 [FREE Full text] [doi: 10.1038/oby.2010.320] [Medline: 21193851]
66. Mathur R, Kim G, Morales W, Sung J, Rooks E, Pokkunuri V, et al. Intestinal Methanobrevibacter smithii but not total bacteria is related to diet-induced weight gain in rats. Obesity (Silver Spring) 2013 Apr 25;21(4):748-754 [FREE Full text] [doi: 10.1002/oby.20277] [Medline: 23712978]
67. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, MetaHIT consortium, et al. Richness of human gut microbiome correlates with metabolic markers. Nature 2013 Aug 29;500(7464):541-546. [doi: 10.1038/nature12506] [Medline: 23985870]
68. Cox LM, Blaser MJ. Antibiotics in early life and obesity. Nat Rev Endocrinol 2015 Mar;11(3):182-190 [FREE Full text] [doi: 10.1038/nrendo.2014.210] [Medline: 25488483]
69. Remely M, Tesar I, Hippe B, Gnauer S, Rust P, Haslberger A. Gut microbiota composition correlates with changes in body fat content due to weight loss. Beneficial Microbes 2015 Aug;6(4):431-439. [doi: 10.3920/bm2014.0104]
70. Jumpertz R, Le DS, Turnbaugh P, Trinidad C, Bogardus C, Gordon JI, et al. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. Am J Clin Nutr 2011 Jul;94(1):58-65 [FREE Full text] [doi: 10.3945/ajcn.110.010132] [Medline: 21543530]
71. Sweeney TE, Morton JM. The human gut microbiome: a review of the effect of obesity and surgically induced weight loss. JAMA Surg 2013 Jun 01;148(6):563-569 [FREE Full text] [doi: 10.1001/jamasurg.2013.5] [Medline: 23571517]
72. Damms-Machado A, Mitra S, Schollenberger AE, Kramer KM, Meile T, Königsrainer A, et al. Effects of surgical and dietary weight loss therapy for obesity on gut microbiota composition and nutrient absorption. Biomed Res Int 2015;2015:1-12 [FREE Full text] [doi: 10.1155/2015/806248] [Medline: 25710027]
73. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr 20;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
74. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
75. Silberman JM, Kaur M, Sletteland J, Venkatesan A. Outcomes in a digital weight management intervention with one-on-one health coaching. PLoS One 2020 Apr 30;15(4):e0232221 [FREE Full text] [doi: 10.1371/journal.pone.0232221] [Medline: 32353035]
76. Bray G, Heisel W, Afshin A, Jensen MD, Dietz WH, Long M, et al. The science of obesity management: an endocrine society scientific statement. Endocr Rev 2018 Apr 01;39(2):79-132 [FREE Full text] [doi: 10.1210/er.2017-00253] [Medline: 29518206]
77. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol 2014 Jul 01;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
78. Harvey-Berino J, West D, Krukowski R, Prewitt E, VanBiervliet A, Ashikaga T, et al. Internet delivered behavioral obesity treatment. Prev Med 2010 Aug;51(2):123-128 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.018] [Medline: 20478333]
79. Tate DF, Wing RR, Winett RA. Using internet technology to deliver a behavioral weight loss program. J Am Med Assoc 2001 Mar 07;285(9):1172-1177. [doi: 10.1001/jama.285.9.1172] [Medline: 11231746]
80. Obino KF, Pereira CA, Lienert RC. Coaching and barriers to weight loss: an integrative review. Diabetes Metab Syndr Obes Targets Ther 2016 Dec;Volume 10:1-11. [doi: 10.2147/dmso.s113874]

XSL•FO
RenderX

81. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 2010 Jul;376(9736):180-188. [doi: 10.1016/S0140-6736(10)60588-0]
82. Kaklamani VG, Sadim M, Hsi A, Offit K, Oddoux C, Ostrer H, et al. Variants of the adiponectin and adiponectin receptor 1 genes and breast cancer risk. Cancer Res 2008 May 01;68(9):3178-3184. [doi: 10.1158/0008-5472.can-08-0533]
83. Clarke R, Halsey J, Bennett D, Lewington S. Homocysteine and vascular disease: review of published results of the homocysteine-lowering trials. J Inherit Metab Dis 2011 Feb 11;34(1):83-91. [doi: 10.1007/s10545-010-9235-y] [Medline: 21069462]
84. Johnson J. β1-adrenergic receptor polymorphisms and antihypertensive response to metoprolol. Clin Pharmacol Ther 2003 Jul;74(1):44-52. [doi: 10.1016/S0009-9236(03)00068-7]
85. Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. PLoS Genet 2007 Nov 9;3(11):e192 [FREE Full text] [doi: 10.1371/journal.pgen.0030192] [Medline: 17997607]
86. van Heel DA, Franke L, Hunt K, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007 Jul;39(7):827-829 [FREE Full text] [doi: 10.1038/ng2058] [Medline: 17558408]
87. CoffeeCaffeine Genetics Consortium, Cornelis MC, Byrne EM, Esko T, Nalls MA, Ganna A, International Parkinson's Disease Genomics Consortium (IPDGC), North American Brain Expression Consortium (NABEC), UK Brain Expression Consortium (UKBEC), et al. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. Mol Psychiatry 2015 May 7;20(5):647-656 [FREE Full text] [doi: 10.1038/mp.2014.107] [Medline: 25288136]
88. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009 Jan 14;41(1):18-24. [doi: 10.1038/ng.274] [Medline: 19079260]
89. ADIPOQ. SNPedia. URL: https://snpedia.com/index.php/ADIPOQ [accessed 2020-10-20]
90. Goyenechea E, Collins L, Parra D, Abete I, Crujeiras A, O'Dell SD, et al. The - 11391 G/A polymorphism of the adiponectin gene promoter is associated with metabolic syndrome traits and the outcome of an energy-restricted diet in obese subjects. Horm Metab Res 2009 Jan 23;41(1):55-61. [doi: 10.1055/s-0028-1087204] [Medline: 18949681]
91. Warodomwichit D, Shen J, Arnett D, Tsai MY, Kabagambe EK, Peacock JM, et al. ADIPOQ polymorphisms, monounsaturated fatty acids, and obesity risk: the GOLDN study. Obesity (Silver Spring) 2009 Mar;17(3):510-517 [FREE Full text] [doi: 10.1038/oby.2008.583] [Medline: 19238139]
92. Gallicchio L, Kalesan B, Huang H, Strickland P, Hoffman SC, Helzlsouer KJ. Genetic polymorphisms of peroxisome proliferator-activated receptors and the risk of cardiovascular morbidity and mortality in a community-based cohort in washington county, Maryland. PPAR Res 2008;2008:1-9 [FREE Full text] [doi: 10.1155/2008/276581] [Medline: 18288282]
93. Leońska-Duniec A, Cieszczyk P, Jastrzębski Z, Jażdżewska A, Lulińska-Kuklik E, Moska W, et al. The polymorphisms of the PPARD gene modify post-training body mass and biochemical parameter changes in women. PLoS One 2018 Aug 29;13(8):e0202557 [FREE Full text] [doi: 10.1371/journal.pone.0202557] [Medline: 30157214]
94. rs4074995. SNPedia. URL: https://www.snpedia.com/index.php/rs4074995 [accessed 2020-10-20]
95. Kestenbaum B, Glazer NL, Köttgen A, Felix JF, Hwang S, Liu Y, et al. Common genetic variants associate with serum phosphorus concentration. J Am Soc Nephrol 2010 Jun 17;21(7):1223-1232. [doi: 10.1681/asn.2009111104]
96. O'Seaghdha CM, Wu H, Yang Q, Kapur K, Guessous I, Zuber AM, SUNLIGHT Consortium, GEFOS Consortium, et al. Meta-analysis of genome-wide association studies identifies six new Loci for serum calcium concentrations. PLoS Genet 2013 Sep 19;9(9):e1003796 [FREE Full text] [doi: 10.1371/journal.pgen.1003796] [Medline: 24068962]
97. 2015-2020 dietary guidelines. Office of Disease Prevention and Health Promotion. 2015. URL: https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines [accessed 2020-10-20]
98. Luo W, Chen F, Guo Z, Wu M, Zhou Z, Yao X. A population association study of PPAR δ gene rs2016520 and rs9794 polymorphisms and haplotypes with body mass index and waist circumference in a Chinese population. Ann Hum Biol 2016 Jun 15;43(1):67-72. [doi: 10.3109/03014460.2015.1023847] [Medline: 26073637]
99. Holzapfel J, Heun R, Lütjohann D, Jessen F, Maier W, Kölsch H. PPARD haplotype influences cholesterol metabolism but is no risk factor of Alzheimer's disease. Neurosci Lett 2006 Nov 06;408(1):57-61. [doi: 10.1016/j.neulet.2006.08.029] [Medline: 16979821]
100. Calcium: fact sheet for health professionals. Office of Dietary Supplements. URL: https://ods.od.nih.gov/factsheets/Calcium-HealthProfessional/ [accessed 2020-10-20]
101. Wu X, Han T, Gao J, Zhang Y, Zhao S, Sun R, et al. Association of serum calcium and insulin resistance with hypertension risk: a prospective population‐based study. J Am Heart Assoc 2019 Jan 08;8(1). [doi: 10.1161/jaha.118.009585]
102. Dalfardi O, Jahandideh D, Omrani G. The correlation of serum calcium level and obesity; is there any explanation? Galen Med J. 2013. URL: https://www.gmj.ir/index.php/gmj/article/view/28/34 [accessed 2020-04-21]
103. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]

XSL•FO
RenderX

104. Kanter R, Caballero B. Global gender disparities in obesity: a review. Adv Nutr 2012 Jul 01;3(4):491-498 [FREE Full text] [doi: 10.3945/an.112.002063] [Medline: 22797984]
105. Lemon SC, Rosal MC, Zapka J, Borg A, Andersen V. Contributions of weight perceptions to weight loss attempts: differences by body mass index and gender. Body Image 2009 Mar;6(2):90-96 [FREE Full text] [doi: 10.1016/j.bodyim.2008.11.004] [Medline: 19188102]
106. Pagoto S, Schneider KL, Oleski JL, Luciani JM, Bodenlos JS, Whited MC. Male inclusion in randomized controlled trials of lifestyle weight loss interventions. Obesity (Silver Spring) 2012 Jun;20(6):1234-1239 [FREE Full text] [doi: 10.1038/oby.2011.140] [Medline: 21633403]
107. Robertson C, Archibald D, Avenell A, Douglas F, Hoddinott P, van Teijlingen E, et al. Systematic reviews of and integrated report on the quantitative, qualitative and economic evidence base for the management of obesity in men. Health Technol Assess 2014 May;18(35):v-vi, xxiii [FREE Full text] [doi: 10.3310/hta18350] [Medline: 24857516]
108. Crane MM, Jeffery RW, Sherwood NE. Exploring gender differences in a randomized trial of weight loss maintenance. Am J Mens Health 2017 Dec;11(2):369-375 [FREE Full text] [doi: 10.1177/1557988316681221] [Medline: 27923968]
109. Gabriele JM, Carpenter BD, Tate DF, Fisher EB. Directive and nondirective e-coach support for weight loss in overweight adults. Ann Behav Med 2011 Apr 25;41(2):252-263 [FREE Full text] [doi: 10.1007/s12160-010-9240-2] [Medline: 21108032]
110. Christensen P, Larsen TM, Westerterp-Plantenga M, Macdonald I, Martinez JA, Handjiev S, et al. Men and women respond differently to rapid weight loss: metabolic outcomes of a multi-centre intervention study after a low-energy diet in 2500 overweight, individuals with pre-diabetes (PREVIEW). Diabetes Obes Metab 2018 Dec 07;20(12):2840-2851 [FREE Full text] [doi: 10.1111/dom.13466] [Medline: 30088336]
111. Celis-Morales C, Livingstone KM, Marsaux CF, Macready AL, Fallaize R, O'Donovan CB, et al. Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial. Int J Epidemiol 2016 Aug 14:dyw186. [doi: 10.1093/ije/dyw186] [Medline: 27524815]
112. Hietaranta-Luoma H, Tahvonen R, Iso-Touru T, Puolijoki H, Hopia A. An intervention study of individual, apoE genotype-based dietary and physical-activity advice: impact on health behavior. J Nutrigenet Nutrigenomics 2014 Feb 20;7(3):161-174. [doi: 10.1159/000371743] [Medline: 25720616]
113. Bloss CS, Wineinger NE, Darst BF, Schork NJ, Topol EJ. Impact of direct-to-consumer genomic testing at long term follow-up. J Med Genet 2013 Jun 04;50(6):393-400. [doi: 10.1136/jmedgenet-2012-101207] [Medline: 23559530]
114. Nielsen DE, El-Sohemy A. Disclosure of genetic information and change in dietary intake: a randomized controlled trial. PLoS One 2014 Nov 14;9(11):e112665. [doi: 10.1371/journal.pone.0112665] [Medline: 25398084]
115. Meisel SF, Beeken RJ, van Jaarsveld CH, Wardle J. Genetic susceptibility testing and readiness to control weight: results from a randomized controlled trial. Obesity (Silver Spring) 2015 Feb 17;23(2):305-312 [FREE Full text] [doi: 10.1002/oby.20958] [Medline: 25522302]
116. Celis-Morales C, Livingstone KM, Marsaux C, Walsh MC, Woolhead C, Forster H, et al. Baseline characteristics of the Food4Me proof of principle study: a web-based randomised controlled trial of personalised nutrition in seven European countries. Proc Nutr Soc 2015 Apr 15;74(OCE1):a [FREE Full text] [doi: 10.1017/s0029665115001330]
117. Freathy RM, Hayes MG, Urbanek M, Lowe LP, Lee H, Ackerman C, HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study: common genetic variants in GCK and TCF7L2 are associated with fasting and postchallenge glucose levels in pregnancy and with the new consensus definition of gestational diabetes mellitus from the International Association of Diabetes and Pregnancy Study Groups. Diabetes 2010 Oct 03;59(10):2682-2689 [FREE Full text] [doi: 10.2337/db10-0177] [Medline: 20682688]
118. Shaat N, Lernmark A, Karlsson E, Ivarsson S, Parikh H, Berntorp K, et al. A variant in the transcription factor 7-like 2 (TCF7L2) gene is associated with an increased risk of gestational diabetes mellitus. Diabetologia 2007 May;50(5):972-979. [doi: 10.1007/s00125-007-0623-2] [Medline: 17342473]
119. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest 2007 Aug;117(8):2155-2163 [FREE Full text] [doi: 10.1172/JCI30706] [Medline: 17671651]
120. Tong Y, Lin Y, Zhang Y, Yang J, Zhang Y, Liu H, et al. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Med Genet 2009 Feb 19;10:15 [FREE Full text] [doi: 10.1186/1471-2350-10-15] [Medline: 19228405]
121. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007 Jun 07;447(7145):661-678 [FREE Full text] [doi: 10.1038/nature05911] [Medline: 17554300]
122. Palmer N, Lehtinen AB, Langefeld CD, Campbell JK, Haffner SM, Norris JM, et al. Association of TCF7L2 gene polymorphisms with reduced acute insulin response in Hispanic Americans. J Clin Endocrinol Metab 2008 Jan;93(1):304-309 [FREE Full text] [doi: 10.1210/jc.2007-1225] [Medline: 17971425]
123. Bhama SCV, Balaji S, Seethalakshmi A. Analysis of the degree of insulin resistance in post menopausal women by using skin temperature measurements and fasting insulin and fasting glucose levels: a case control study. J Clin Diagn Res 2012:1644-1647 [FREE Full text] [doi: 10.7860/jcdr/2012/4377.2646]

XSL•FO
RenderX

124. Heianza Y, Qi L. Gene-diet interaction and precision nutrition in obesity. Int J Mol Sci 2017 Apr 07;18(4):787 [FREE Full text] [doi: 10.3390/ijms18040787] [Medline: 28387720]

Abbreviations

FTO: alpha-ketoglutarate-dependent dioxygenase
LCT: lactase
MUFA: monounsaturated fat
RCT: randomized controlled trial
SNP: single-nucleotide polymorphism

Edited by G Eysenbach; submitted 01.11.20; peer-reviewed by S Rostam Niakan Kalhori, J Alvarez Pitti, AV Das; comments to author 14.12.20; revised version received 18.12.20; accepted 11.04.21; published 19.05.21.

Please cite as:
Sinha R, Kachru D, Ricchetti RR, Singh-Rambiritch S, Muthukumar KM, Singaravel V, Irudayanathan C, Reddy-Sinha C, Junaid I, Sharma G, Francis-Lyon PA
Leveraging Genomic Associations in Precision Digital Care for Weight Loss: Cohort Study
J Med Internet Res 2021;23(5):e25401
URL: https://www.jmir.org/2021/5/e25401
doi:10.2196/25401
PMID:33849843

©Ranjan Sinha, Dashyanng Kachru, Roshni Ray Ricchetti, Simitha Singh-Rambiritch, Karthik Marimuthu Muthukumar, Vidhya Singaravel, Carmel Irudayanathan, Chandana Reddy-Sinha, Imran Junaid, Garima Sharma, Patricia Alice Francis-Lyon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Letter to the Editor

COVID-19 as a "Force Majeure" for Non–COVID-19 Clinical and Translational Research. Comment on "Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study"

Petar Milovanovic[1], MD, PhD; Igor Dumic[2], MD

[1]Faculty of Medicine, University of Belgrade, Belgrade

[2]Division of Hospital Medicine, Mayo Clinic Health System, Eau Claire, WI, United States

Corresponding Author:
Petar Milovanovic, MD, PhD
Faculty of Medicine
University of Belgrade
Dr Subotica 4/2
Belgrade, 11000
Phone: 381 66413225
Email: drpmilovanovic@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2020/11/e21559

Comment in: https://www.jmir.org/2021/5/e29156/

(J Med Internet Res 2021;23(5):e27937) doi:10.2196/27937

KEYWORDS

COVID-19; SARS-CoV-2; coronavirus; pandemic; topic modeling; research; literature; medical research; publishing; force majeure

We read with interest the recent article on medical publication of COVID-19-related research by Älgå et al [1]. We acknowledge their topic modeling approach to analyze publications about COVID-19. However, we would like to emphasize that although research on COVID-19 is receiving wide support, other areas in medicine are facing a lasting problem. Many studies are blocked or slowed down due to delayed or interrupted recruitment of patients and/or delayed analyses or processing. With entire wards and hospitals converted to COVID-19 units, the numbers of elective surgeries and hospitalizations are substantially reduced and many nonessential hospital visits are postponed, substantially reducing the possibilities for patient recruitment and follow-up. Many resources primarily used for research and clinical and translational studies have shifted to the diagnostics and management of COVID-19; although this shift testifies to the positive robustness and flexibility of the medical system, it inevitably slows down other, at this moment "nonessential," research activities. Many physicians are working on COVID-19–related topics and with patients with COVID-19, temporarily suspending their regular research activities. In the situation of lockdown set by governments, researchers continued working on the already available specimens and data from their home offices and communicating via web-based platforms [2]. Despite no apparent decline in manuscript submissions, a reduction in publishing by women has already been observed in some fields [3], suggesting that the real consequences for scientific publishing are yet to come. Processing already available data and manuscript preparation is possible from home, but patient recruitment and conduction of ongoing clinical trials are essential for further scientific breakthroughs. In this context, scientific progress will be one of many victims of COVID-19, and discoveries related to a spectrum of diseases other than COVID-19 will be delayed, especially if funding is reduced [4,5]. Although risk management strategies are a regular part of scientific work and are already obligatory at the step of writing grant proposals, this prolonged uncertainty and worldwide blockade go beyond what anyone could have imagined as a "force majeure." Therefore, we need to find new, creative solutions for continuing research in such situations, guaranteeing safety for patients, researchers, and research quality, to ensure that the society is not deprived of new discoveries in medicine. In this context, computer-based

XSL•FO
RenderX

simulations and analysis of open data will become increasingly important in the future.

As reported by Älgå et al [1], the types and topics of published articles on COVID-19 have evolved during the pandemic, updating the scientific audience in a timely fashion. In contrast, in some previous epidemics, such as H1N1 [1] or Middle East respiratory syndrome and severe acute respiratory syndrome, most results were published after the outbreak [6]. From the administrative side, we think that the COVID-19 pandemic should be accepted as a "force majeure" for non–COVID-19 research, where the funding agencies should provide flexibility and allow extensions or delays of currently running grants. Indeed, major funding agencies already announced supportive measures, such as "no-cost extensions" and reallocation of funds [7], but these are usually granted on a case-by-case basis [8]. However, if COVID-19 affects the realization of a grant at the most sensitive phases (eg, during patient recruitment), the delivery of the results is seriously endangered, even when the funding agencies allow time extensions. From the scientific side, we recommend the following solutions wherever applicable. First, as a *damage-minimizing strategy* for projects that had been already running when the pandemic started, the researchers may try to reevaluate their goals (with consent from the funder) and opt for realistic targets if the original endpoints are no longer achievable. In translational science, when a particular technique is unavailable due to reallocation for COVID-19 patient care, closed laboratories, or restricted international cooperation, alternative but available techniques should be considered. As the situation gets better, researchers could complement the research agenda with the originally planned materials and methods. However, for new grant proposals that are about to be submitted, it is mandatory to provide a clear plan of implementation that would take COVID-19 into account. Specifically, the implementation strategy may provide two scenarios (one more optimistic and one more pessimistic) with respect to the unpredictable times. In any case, a list of measures for ensuring the safety of patients and researchers along with as realistic an estimation as possible of the number of patients and tests required and prepared alternatives for various research methods may increase the confidence of both the researchers and the funders in realistic and good implementation. In this context, we believe that use of the internet and internet-based technologies should be encouraged. Indeed, medical research should not stop, but adaptations to new conditions are crucial. Although we hope that vaccination will soon bring the pandemic to an end, fear of new and mutated strains of the same or different viruses and a rebound pandemic or some new pandemics should keep us thinking about how to do research in such no-longer-so-unimaginable situations.

Acknowledgments

PM thanks the Science Fund of the Republic of Serbia (PROMIS, DiaBoNet, 6064549).

Authors' Contributions

Both authors contributed to writing the paper.

Conflicts of Interest

None declared.

References

1. Älgå A, Eriksson O, Nordberg M. Analysis of scientific publications during the early phase of the COVID-19 pandemic: topic modeling study. J Med Internet Res 2020 Nov 10;22(11):e21559 [FREE Full text] [doi: 10.2196/21559] [Medline: 33031049]
2. Korbel JO, Stegle O. Effects of the COVID-19 pandemic on life scientists. Genome Biol 2020 May 11;21(1):113-141 [FREE Full text] [doi: 10.1186/s13059-020-02031-1] [Medline: 32393316]
3. Viglione G. Are women publishing less during the pandemic? Here's what the data say. Nature 2020 May;581(7809):365-366. [doi: 10.1038/d41586-020-01294-9] [Medline: 32433639]
4. Weiner DL, Balasubramaniam V, Shah SI, Javier JR, Pediatric Policy Council. COVID-19 impact on research, lessons learned from COVID-19 research, implications for pediatric research. Pediatr Res 2020 Aug;88(2):148-150. [doi: 10.1038/s41390-020-1006-3] [Medline: 32544922]
5. Peeters A, Mullins G, Becker D, Orellana L, Livingston P. COVID-19's impact on Australia's health research workforce. Lancet 2020 Aug;396(10249):461. [doi: 10.1016/S0140-6736(20)31533-6]
6. Takagi H. COVID-19 pandemic and medical publication. Eur J Clin Invest 2020 Nov 01;50(11):e13394-e13341. [doi: 10.1111/eci.13394] [Medline: 32886801]
7. Stoye E. How research funders are tackling coronavirus disruption. Nature 2020 Apr 17. [doi: 10.1038/d41586-020-01120-2] [Medline: 32303733]
8. Can the clause on "force majeure" in the Horizon 2020 grant agreement be invoked in relation to the COVID-19 outbreak? European Commission. 2020 Mar 24. URL: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/12945 [accessed 2021-05-17]

XSL•FO
RenderX

Edited by T Derrick; submitted 13.02.21; this is a non–peer-reviewed article;accepted 13.05.21; published 20.05.21.

Please cite as:
Milovanovic P, Dumic I
COVID-19 as a "Force Majeure" for Non–COVID-19 Clinical and Translational Research. Comment on "Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study"
J Med Internet Res 2021;23(5):e27937
URL: https://www.jmir.org/2021/5/e27937
doi:10.2196/27937
PMID:33989167

©Petar Milovanovic, Igor Dumic. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Letter to the Editor

Authors' Reply to: COVID-19 as a “Force Majeure” for Non–COVID-19 Clinical and Translational Research. Comment on “Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study”

Andreas Älgå[1,2], MD, PhD; Oskar Eriksson[3], BSc; Martin Nordberg[1], MD, PhD

[1]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

[2]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

[3]DataRobot Inc, Stockholm, Sweden

Corresponding Author:
Andreas Älgå, MD, PhD
Department of Clinical Science and Education, Södersjukhuset
Karolinska Institutet
Sjukhusbacken 10
Stockholm, 11883
Sweden
Phone: 46 86161000
Email: andreas.alga@ki.se

Related Articles:

Comment on: https://www.jmir.org/2020/11/e21559

Comment on: https://www.jmir.org/2021/5/e27937

(J Med Internet Res 2021;23(5):e29156) doi:10.2196/29156

KEYWORDS

COVID-19; SARS-CoV-2; coronavirus; pandemic; topic modeling; research; literature; medical research; publishing; force majeure

We thank Petar Milovanovic and Igor Dumic [1] for raising the complex issue of challenges associated with non–COVID-19–related research. We share their view that many of the effects that the COVID-19 pandemic has had on such research are likely yet to be seen and that researchers and funders need to find pragmatic ways to deal with this new reality. However, some of the authors' suggested actions, such as to adjust endpoints and power calculations, risk compromising research quality and should be done with caution. In times and settings associated with poor preconditions for clinical studies, we encourage researchers to assess the utility of alternative data sources, such as readily available medical image and routine clinical data.

We concur with the authors that analysis of open data and computer-based simulations have the potential to play a vital role in the future of medical research, apart from experimental and clinical studies. Additionally, we believe that methods like the one used in our study [2] may also prove to be a way forward. The number of published scientific studies is rapidly increasing, and traditional ways of compiling study results to generate new knowledge are slow and not suitable for these numbers. Therefore, the application of existing machine learning text analysis methods and the development of new ones holds promise of scientific gains not only from *existing data* but also from *already completed studies*.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Milovanovic P, Dumic I. COVID-19 as a “Force Majeure” for Non–COVID-19 Clinical and Translational Research. Comment on "Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study". J Med Internet Res 2021 May [FREE Full text] [doi: 10.2196/27937] [Medline: 33989167]
2. Älgå A, Eriksson O, Nordberg M. Analysis of scientific publications during the early phase of the COVID-19 pandemic: topic modeling study. J Med Internet Res 2020 Nov 10;22(11):e21559 [FREE Full text] [doi: 10.2196/21559] [Medline: 33031049]

Edited by T Derrick; submitted 01.04.21; this is a non–peer-reviewed article;accepted 13.05.21; published 20.05.21.

Please cite as:
Älgå A, Eriksson O, Nordberg M
Authors' Reply to: COVID-19 as a “Force Majeure” for Non–COVID-19 Clinical and Translational Research. Comment on “Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study”
J Med Internet Res 2021;23(5):e29156
URL: https://www.jmir.org/2021/5/e29156
doi:10.2196/29156
PMID:33989170

©Andreas Älgå, Oskar Eriksson, Martin Nordberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Letter to the Editor

Methodological Clarifications and Generalizing From Weibo Data. Comment on “Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo”

Om Prakash Yadav[1], MDS

Division of Community Health and Humanities, Memorial University of Newfoundland, St John's, NL, Canada

Corresponding Author:
Om Prakash Yadav, MDS
Division of Community Health and Humanities
Memorial University of Newfoundland
230 Elizabeth Ave
St John's, NL, A1C 5S7
Canada
Phone: 1 7097706592
Email: opyadav@mun.ca

Related Articles:

Comment on: https://www.jmir.org/2020/6/e19981/

Comment in: https://www.jmir.org/2021/5/e29145/

(J Med Internet Res 2021;23(5):e26255) doi:10.2196/26255

KEYWORDS

COVID-19; dentistry; oral health; dental health; online health; social media; tweet; Weibo; China; health information

I read the article regarding the nature and diffusion of COVID-19–related oral health information conducted by Tao et al [1] with great interest. Information, education, and communication are the best tools to prevent and control any disease. The paper described the temporal and geographic distribution of COVID-19–related oral health information circulated on the social media platform Weibo in China. The study’s focus was to assess COVID-19–related oral health information and explore public reaction.

The introduction was built on existing literature describing COVID-19, its impact on dental practice and specific concerns, and the necessity of information on social media. The authors presented the need for the study in this section; however, it could be elaborated to better understand the target audience.

The authors implemented their research with reproducible methods and described all necessary steps, including screening criteria, piloting and reliability assessment, and statistical methods. The authors included 4 search keywords (“pneumonia of unknown cause,” “coronavirus,” “COVID-19,” and “epidemic”) for COVID-19; however, only 2 keywords (“stomatology” and “dentistry”) were selected to search for oral health and disease-related information. The authors could have included at least 2 additional popular keywords (“oral examination” and “oral health”) to make the search more comprehensive [2,3]. The authors included 15,900 tweets, which strengthened the precision and power of the study.

The study results were well described and presented in a logical and comprehensible manner using tables and graphs. However, some inconsistencies were observed in the x-axis parameters in some figures, and one typo was observed in one of the tables’ data. The authors could also have highlighted the point related to the supply and demand for dental health services during the epidemic.

The authors mentioned some limitations, such as the Weibo platform being more popular among younger age groups than older ones, the nonavailability of information regarding individual users’ characteristics, and the nonavailability of information on online consultation quality services. However, users’ education, academic background, job profile, socioeconomic status, past dental treatment experience, and status of existing oral health problems are essential to categorize and associate oral health information within different parameters.

XSL•FO
RenderX

Other limitations include the use of a single social media source and data collection within a short period of time. These limitations can affect the generalizability of the study. Despite the discussed limitations, the authors are to be congratulated for their novel enrichment of evidence through this exploratory research, and with great excitement, I anticipate future research to highlight further circulation of COVID-19–related oral information carried out in a broader time frame and extended to various geographical locations.

Conflicts of Interest

None declared.

References

1. Tao ZY, Chu G, McGrath C, Hua F, Leung YY, Yang WF, et al. Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]
2. Park JE, Cho JW, Jang JH. Keyword Trends for Mother-Child Oral Health in Korea Based on Social Media Big Data from Naver. Healthc Inform Res 2020 Jul;26(3):212-219 [FREE Full text] [doi: 10.4258/hir.2020.26.3.212] [Medline: 32819039]
3. LaLonde K. COVID-19 Literature Searches. Medical Library Association. 2020 Apr 9. URL: https://www.mlanet.org/page/covid-19-literature-searching [accessed 2020-12-03]

Edited by T Derrick; submitted 03.12.20; peer-reviewed by YX Su; accepted 13.05.21; published 21.05.21.

Please cite as:
Yadav OP
Methodological Clarifications and Generalizing From Weibo Data. Comment on "Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo"
J Med Internet Res 2021;23(5):e26255
URL: https://www.jmir.org/2021/5/e26255
doi:10.2196/26255
PMID:33989161

©Om Prakash Yadav. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.05.2021. 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

Letter to the Editor

Authors' Reply to: Methodological Clarifications and Generalizing From Weibo Data. Comment on "Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo"

Zhuo-Ying Tao[1], BDS, MDS; Yu-Xiong Su[1], BDS, MD, PhD

Division of Oral and Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong, Hong Kong (China)

Corresponding Author:
Yu-Xiong Su, BDS, MD, PhD
Division of Oral and Maxillofacial Surgery
Faculty of Dentistry
The University of Hong Kong
34 Hospital Road, Sai Ying Pun
Hong Kong
Hong Kong (China)
Phone: 852 2859 0267
Email: richsu@hku.hk

Related Articles:

Comment on: https://www.jmir.org/2021/5/e26255/

Comment on: http://www.jmir.org/2020/6/e19981/

(J Med Internet Res 2021;23(5):e29145) doi:10.2196/29145

KEYWORDS

COVID-19; dentistry; oral health; dental health; online health; social media; tweet; Weibo; China; health information

We sincerely thank Yadav [1] for the comments on our paper, "Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo" [2]. We appreciate the insightful and constructive opinions on this topic. Indeed, oral health information on social media during the pandemic is important and worthy of in-depth study. Here, we would like to further discuss this topic.

The aim of our study was to investigate oral health information on Chinese social media during the COVID-19 outbreak. As a scientific publication with open access, this study is available to any member of the general public who is interested in this topic; thus, our paper is not oriented toward a specific audience. In this study, the nature and diffusion of COVID-19–related oral health information on Weibo was overviewed. The government can get useful ideas for public health surveillance or policymaking during the epidemic, while dental care providers and potential patients might be inspired knowing that social media is a good communication tool for exchanging oral health knowledge and dental care needs during an emerging crisis.

In terms of the keywords used in this study, we only selected "stomatology" and "dentistry" because these 2 keywords in Chinese provide broad coverage, including oral, mouth, tooth, etc. As such, when we conducted the search with "stomatology" and "dentistry," tweets containing "oral examination" or "oral health" were included in the results as well.

As pointed out by Yadav [1], there was inconsistency in the x-axis parameters (date) in Figure 5 in our published manuscript [2]. This was due to the difference in the starting point of the x-axis, which was the date when each piece of misinformation began to be disseminated on Weibo, while the ending point of the x-axis was March 16, 1 day before we carried out the search. Additionally, there was a typo in Table 2, in which "prosthetics" should be replaced by "prostheses."

We have discussed the imbalance in dental service supply-demand during the epidemic and proposed that social media might be a potential solution at this unprecedented time. The tweets on Weibo reflected the great needs and short supply of dental treatment, but the specific fluctuation of dental demand-supply relationships with time was not able to be evaluated through Weibo. Nevertheless, social media can still serve as an important market indicator for both government policymakers and commercial companies to meet oral health needs during the pandemic.

XSL•FO
RenderX

Obviously, oral health information tweeted on social media is significantly influenced by the users' socioeconomic status (educational background, job profile, income level) and oral health status (dental treatment experience, existing oral health problems). However, when performing social media–based studies, researchers have no access to this information due to the confidentiality and privacy of social media users, which means it is technically impossible to estimate the association between users' background and the information they tweet about. This can be considered as an innate limitation of social media studies.

In China, Weibo is representative of all social media, especially during emerging epidemics when active users and tweets increase sharply [3,4]. According to the financial report of Sina Weibo for the first quarter of 2020 [5], the number of active users reached as high as 550 million per month and increased by 85 million compared with the same period last year. Therefore, we chose Weibo as the only social media platform for our study to avoid potential overlapping of users registered on other platforms. The large amount of Weibo users provided sufficient data for conducting the study.

The outbreak of COVID-19 in Wuhan, China, started on December 31, 2019. Since mid-March 2020, COVID-19 has been well controlled in Wuhan, with less than 10 new cases per day. Since then, dental hospitals and clinics have started to provide services to the public, leading to a significant reduction in tweets about COVID-19–related dentistry information. Accordingly, the scenario in Wuhan provides a unique and important social model to study the pandemic and related oral health information within a short time frame. Our study, covering the time period from December 31, 2019, to March 16, 2020, completely documents the changes in oral health information on social media from the beginning to the end of the outbreak in the world's first COVID-19 epicenter.

The distribution of oral health information is closely related to the health care system of individual countries. We agree that future studies should be carried out to highlight further circulation of COVID-19–related oral information using a broader time frame and extended geographical locations. Comparisons of information between different geographical locations and different social media platforms in various languages will be of great value. It will not only provide insight into the global influence of the pandemic but will also shed light on the policy-making of health care systems around the world. This highlights a truly interesting and impactful research direction for the future.

Conflicts of Interest

None declared.

References

1. Yadav OP. Methodological Clarifications and Generalizing From Weibo Data. Comment on "Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo". J Med Internet Res 2021 May;22(5):e26555 [FREE Full text] [doi: 10.2196/26255] [Medline: 33989161]
2. Tao ZY, Chu G, McGrath C, Hua F, Leung YY, Yang WF, et al. Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]
3. Fung ICH, Fu KW, Ying Y, Schaible B, Hao Y, Chan CH, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013 Dec 20;2(1):31 [FREE Full text] [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]
4. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
5. Weibo monthly active users reach 550 million, revenue surpasses Wall Street expectations. Sina Finance. 2020 May 19. URL: https://finance.sina.com.cn/stock/usstock/c/2020-05-19/doc-iircuyvi3963989.shtml [accessed 2021-05-17]

Edited by T Derrick; submitted 27.03.21; this is a non–peer-reviewed article;accepted 13.05.21; published 21.05.21.

Please cite as:
Tao ZY, Su YX
Authors' Reply to: Methodological Clarifications and Generalizing From Weibo Data. Comment on "Nature and Diffusion of COVID-19–related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo"
J Med Internet Res 2021;23(5):e29145
URL: https://www.jmir.org/2021/5/e29145
doi:10.2196/29145
PMID:33989166

XSL•FO
RenderX

©Zhuo-Ying Tao, Yu-Xiong Su. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.05.2021. 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

Letter to the Editor

Screening Tools: Their Intended Audiences and Purposes. Comment on “Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study”

Elizabeth Millen[1], BSc; Andreas Gilsdorf[1], MD, PhD; Matthew Fenech[1], MD, PhD; Stephen Gilbert[1], PhD

Ada Health, Berlin, Germany

Corresponding Author:
Elizabeth Millen, BSc
Ada Health
Karl-Liebknecht-Str 1
Berlin, 10178
Germany
Phone: 49 01520713083
Email: elizabeth.millen@ada.com

Related Articles:

Comment on: https://www.jmir.org/2020/10/e21299/

Comment in: https://www.jmir.org/2021/5/e26543

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e31268/

(J Med Internet Res 2021;23(5):e26148) doi:10.2196/26148

KEYWORDS

COVID-19; symptom checkers; benchmark; digital health; symptom; chatbot; accuracy

We congratulate Munsch et al [1] on their recent publication, “Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study.” The study investigated the relative performance of web-based COVID-19 screening tools between April 3 and 9, 2020. It is important to have literature comparing the relative performance of tools; however, it is also important that the approaches used in reports of this type compare “like with like” and are reported rigorously.

There are a number of important limitations in Munsch et al’s [1] paper, the first being that the principal results reported are not fully addressed in the publication, and the second being it is an inappropriate evaluation of COVID-19 screeners designed for use by laypersons at home. COVID-19 screeners were brought into the market at a time of need to lower the burden on health authorities and health care institutions [2]. Overall, they were meant for use at home by people who were either worried about their symptoms or who were anxious about the pandemic in general and in need of reassurance. These people could either be put at ease that their symptoms were likely not caused by COVID-19, or be provided with appropriate localized advice through screeners tailored to their region.

The paper [1] does not investigate the COVID-19 screeners according to their intended purpose and compares tools with fundamentally different intended purposes. The Symptoma tool, developed by the paper’s authors, was designed to use the professional interpretation of both patients’ symptoms and the results of diagnostic tests, such as interpreted CT (computed tomography) images (eg, ground-glass opacities); both provide a predictive likelihood of COVID-19 status. All the other screeners examined in the study were designed for use at home by layperson users, providing a first line of advice to those experiencing possible COVID-19 symptoms. Importantly, none of these tools provide a diagnosis, neither their design concept nor regulatory approval includes this, and testing their diagnostic accuracy is therefore testing something they are specifically labeled as not providing. A meaningful evaluation would be to evaluate the appropriateness of the advice given according to the use context of these screeners (ie, do the screeners provide appropriate information and advice to users based on their symptoms, risk factors, exposure, and their location [including local COVID-19 status and national guidelines]?).

The ability of Symptoma to make use of the professional interpretation of clinical findings might be useful in some hospital settings as a clinical decision support tool, but COVID-19 tests are available in this setting, and at this stage in the patient journey, the tool cannot lessen the burden on the

XSL•FO
RenderX

health care system. Based on the arguments above, it is critical to interpret the results in Multimedia Appendices 8-12 [1], which show that Symptoma is not superior in specificity and sensitivity as claimed in the main manuscript, and to remember that the appropriateness of specific advice in a given situation is more important to the user than either specificity or sensitivity.

Conflicts of Interest

All authors are employees of Ada Health GmbH.

References

1. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study. J Med Internet Res 2020 Oct 06;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
2. Ross C. Symptom checker steers Covid-19 patients to care. STAT. 2021 Apr 21. URL: https://www.statnews.com/2020/04/21/coronavirus-symptom-checker-steers-patients-to-care-far-away/ [accessed 2020-07-22]

Abbreviations

CT: computed tomography

Edited by T Derrick; submitted 30.11.20; peer-reviewed by B Knapp; accepted 13.05.21; published 21.05.21.

Please cite as:
Millen E, Gilsdorf A, Fenech M, Gilbert S
Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study"
J Med Internet Res 2021;23(5):e26148
URL: https://www.jmir.org/2021/5/e26148
doi:10.2196/26148
PMID:33989169

©Elizabeth Millen, Andreas Gilsdorf, Matthew Fenech, Stephen Gilbert. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.05.2021. 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

Letter to the Editor

Authors' Reply to: Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study"

Nicolas Munsch[1], MSc; Alistair Martin[1], DPhil; Stefanie Gruarin[2], MD, PhD; Jama Nateqi[2,3], MD; Isselmou Abdarahmane[1], MSc; Rafael Weingartner-Ortner[1,2], MSc; Bernhard Knapp[1,4], PhD

[1]Data Science Department, Symptoma, Vienna, Austria

[2]Medical Department, Symptoma, Attersee, Austria

[3]Department of Internal Medicine, Paracelsus Medical University, Salzburg, Austria

[4]Department of Computer Science, University of Applied Sciences – Technikum Wien, Vienna, Austria

Corresponding Author:
Bernhard Knapp, PhD
Data Science Department
Symptoma
Kundmanngasse 21
Vienna, 1030
Austria
Phone: 43 662458206
Email: science@symptoma.com

Related Articles:

Comment on: https://www.jmir.org/2020/10/e21299/

Comment on: https://www.jmir.org/2021/5/e26148

(J Med Internet Res 2021;23(5):e26543) doi:10.2196/26543

KEYWORDS

COVID-19; symptom checkers; benchmark; digital health; symptom; chatbot; accuracy

We thank Millen et al [1] (Ada Health) for having taken the time to read our recently published COVID-19 symptom checker comparison study [2] and for their letter to the editor. Millen et al [1] identified three opportunities for improvement to our manuscript. We wish to provide an itemized response to their letter as outlined below.

First, Millen et al [1] state that the Symptoma symptom checker is a fundamentally different tool, which "might be useful in some hospital settings as a clinical decision support tool" but should not be compared to other layperson COVID-19 symptom checkers. We disagree with this assessment. Indeed, the opposite is true: the same Symptoma engine is widely used by laypersons and therefore the comparison as executed in our publication is appropriate.

Second, Millen et al [1] state that Symptoma's superior accuracy is due to its unique ability "to make use of professional interpretation of clinical findings" and that such data should not be used by any of the symptom checkers. At Symptoma, we have found that the greater the wealth of data flowing into the symptom checker AI (artificial intelligence), the better the output and results generated for our users. Further, Millen et al [1] state that "Symptoma does not indeed perform superiorly" if no clinical findings are taken into account while referring to the appendix of our study [2]. We respectfully disagree with this on several fronts: (1) the appendix shows a different analysis than that referred to by Millen et al [1] and (2) that if the analysis is modified to fit Millen et al's [1] suggestion, which we consider less appropriate, the ranking still remains unchanged. The F1 scores will be slightly modified (from 0.92 to 0.90 for "high risk" only while it remains at 0.91 for "high risk" and "medium risk") while still ranking Symptoma first.

Lastly, Millen et al [1] suggest that the evaluation of their symptom checker should not be based on "accuracy." The most prevalent outcome parameter of comparative symptom checker studies is diagnostic accuracy [3-5]. Even though the symptom checker results do not represent a diagnosis, the importance of accurate results as generated by a symptom checker is paramount to its functionality. Other outcome parameters such as the ones suggested by Millen et al [1] are poorly measurable, harbor the potential for bias, and lack comparability.

XSL•FO
RenderX

Conflicts of Interest

All authors are employees of Symptoma GmbH. JN holds shares in Symptoma.

References

1. Millen E, Gilsdorf A, Fenech M, Gilbert S. Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study". J Med Internet Res 2021 May;23(5):e26148 [FREE Full text] [doi: 10.2196/26148]
2. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study. J Med Internet Res 2020 Oct 06;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
3. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
4. Nateqi J, Lin S, Krobath H, Gruarin S, Lutz T, Dvorak T, et al. [From symptom to diagnosis-symptom checkers re-evaluated : Are symptom checkers finally sufficient and accurate to use? An update from the ENT perspective]. HNO 2019 May;67(5):334-342. [doi: 10.1007/s00106-019-0666-y] [Medline: 30993374]
5. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of Physician and Computer Diagnostic Accuracy. JAMA Intern Med 2016 Dec 01;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]

Abbreviations

AI: artificial intelligence

Edited by T Derrick; submitted 16.12.20; this is a non–peer-reviewed article;accepted 13.05.21; published 21.05.21.

Please cite as:
Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, Knapp B
Authors' Reply to: Screening Tools: Their Intended Audiences and Purposes. Comment on "Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study"
J Med Internet Res 2021;23(5):e26543
URL: https://www.jmir.org/2021/5/e26543
doi:10.2196/26543
PMID:33989162

©Nicolas Munsch, Alistair Martin, Stefanie Gruarin, Jama Nateqi, Isselmou Abdarahmane, Rafael Weingartner-Ortner, Bernhard Knapp. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.05.2021. 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

Letter to the Editor

Minimizing Selection and Classification Biases. Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing"

Francisco Martos Pérez[1], MD, PhD; Ricardo Gomez Huelgas[2], MD, PhD; María Dolores Martín Escalante[1], MD, PhD; José Manuel Casas Rojo[3], MD

[1]Department of Internal Medicine, Hospital Costa del Sol, Marbella, Spain

[2]Internal Medicine Department, Hospital Regional de Málaga, Málaga, Spain

[3]Internal Medicine Department, Infanta Cristina University Hospital, Parla (Madrid), Spain

Corresponding Author:
Francisco Martos Pérez, MD, PhD
Department of Internal Medicine
Hospital Costa del Sol
Autovía A-7, Km 187
Marbella, 29603
Spain
Phone: 34 658927715
Email: pacomartos1@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2020/10/e21801/

Comment in: https://www.jmir.org/2021/5/e29405/

(J Med Internet Res 2021;23(5):e27142) doi:10.2196/27142

KEYWORDS

artificial intelligence; big data; COVID-19; electronic health records; tachypnea; SARS-CoV-2; predictive model; prognosis; classification bias; critical care

The paper by Izquierdo et al [1], published in the recent issue of the *Journal of Medical Internet Research*, employed a combination of conventional and machine learning tools to describe the clinical characteristics of patients with COVID-19 and the factors that predict intensive care unit (ICU) admission. We would like to make some comments about its design.

The authors should have provided the proportion of patients with a positive microbiological diagnosis. If the artificial intelligence software's capacity for retrieving this information is limited in some way, this should be explained. The classification bias introduced by the lack of microbiological confirmation may have been significant since the study includes patients from January 1, 2020. Although some undiagnosed cases have likely been present prior to the first declared case (March 1, 2020) [2] in Castilla-La Mancha, it is improbable that there were many of them.

ICU admissions are related to many factors not addressed in the study. The decision not to admit a patient to the ICU because of short life expectancy, low quality of life, or high burden of comorbidities may have had a great impact during the first wave of the COVID-19 pandemic, when a scarcity of ICU beds manifested in some regions of Spain. The 6.1% ICU admission rate reported by the authors was 36% lower than the 8.3% reported in a national survey of 15,111 patients from 150 hospitals in Spain [3]. We could hypothesize that the patients included in the study had a milder form of the disease. However, given the absence of a microbiological diagnosis in an unknown percentage of patients, the inclusion of a significant proportion of patients without a real COVID-19 diagnosis cannot be ruled out. These doubts could have been resolved if a microbiological diagnosis had been a requisite for inclusion. The mortality rate, the most robust and relevant endpoint, should also have been reported and the factors related to it analyzed.

XSL•FO
RenderX

Artificial intelligence is having an increasing impact on the rate of health care information processing. However, minimization of selection and classification biases should be guaranteed in the design of investigations. In this case, this could have been achieved by including only microbiologically confirmed cases and prolonging the period of inclusion, since most COVID-19 cases emerged after the end date of the study inclusion period. These changes in the design would have allowed for a better evaluation of the performance of artificial intelligence techniques, making the results obtained in the sample closer to those of the real population.

Conflicts of Interest

None declared.

References

1. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. J Med Internet Res 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: 10.2196/21801] [Medline: 33090964]
2. Europa Press. 2020 Mar 1. URL: https://www.europapress.es/castilla-lamancha/noticia-varon-62-anos-ingresado-guadalajara-primer-caso-positivo-coronavirus-lm-20200301103741.html [accessed 2020-01-08]
3. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. Rev Clin Esp (Barc) 2020 Nov;220(8):480-494 [FREE Full text] [doi: 10.1016/j.rce.2020.07.003] [Medline: 32762922]

Abbreviations

ICU: intensive care unit

Edited by T Derrick; submitted 12.01.21; peer-reviewed by M Elbattah, J Marconato, T Capote; accepted 13.05.21; published 26.05.21.

Please cite as:
Martos Pérez F, Gomez Huelgas R, Martín Escalante MD, Casas Rojo JM
Minimizing Selection and Classification Biases. Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing"
J Med Internet Res 2021;23(5):e27142
URL: https://www.jmir.org/2021/5/e27142
doi:10.2196/27142
PMID:33989163

©Francisco Martos Pérez, Ricardo Gomez Huelgas, María Dolores Martín Escalante, José Manuel Casas Rojo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2021. 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

Letter to the Editor

Authors' Reply to: Minimizing Selection and Classification Biases Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing"

Jose Luis Izquierdo[1], MD; Joan B Soriano[2], MD

[1]Universidad de Alcalá, Madrid, Spain

[2]Hospital Universitario de La Princesa, Madrid, Spain

Corresponding Author:
Joan B Soriano, MD
Hospital Universitario de La Princesa
Diego de Leon 62
Servicio de Neumología
Madrid, 28006
Spain
Email: jbsoriano2@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2020/10/e21801/

Comment on: https://www.jmir.org/2021/5/e27142

(J Med Internet Res 2021;23(5):e29405) doi:10.2196/29405

KEYWORDS

artificial intelligence; big data; COVID-19; electronic health records; tachypnea; SARS-CoV-2; predictive model; prognosis; classification bias; critical care

We acknowledge the letter by Martos Pérez et al [1] and take this opportunity to clarify related issues from our publication [2]. Of our 10,504 patients with COVID-19, 2737 (26.5%) were tested with PCR (polymerase chain reaction). Within the 5 provinces of Castilla-La Mancha, the province that tested the most was Toledo (28.9%), while the least was Guadalajara (21.2%). Those patients in whom PCR was performed were 6.5 years older (63.0 vs 56.5 years). All these differences were highly statistically significant.

You must take into account that our study period was from March 1 to 29, 2020, and including only microbiologically confirmed cases or prolonging the period of inclusion would have resulted in a biased assessment. From March 30, 2020, onwards, most intensive care units (ICUs) at our hospitals collapsed and ICU admissions were highly distorted due to a lack of beds. As we commented in the *Discussion* section, the ICU capacity in Castilla-La Mancha during the study period had not yet been compromised, which protects against possible bias in our training data (all patients requiring critical care were indeed admitted to the ICU). Therefore, it is unlikely that the absence of a confirmed diagnosis with PCR during the first weeks of the pandemic influenced our results. This was a generalized situation throughout Spain and in most European countries early in 2020. At that time, when a patient was hospitalized, a wide battery of viruses was considered for which there were reagents before performing PCR for coronaviruses. Patients seen during the month of March, in the midst of an avalanche of COVID-19 cases in our region, with negative tests for other viruses and clinical, radiologic, and blood tests highly compatible, did not raise doubts about their diagnosis of COVID-19, and the probability of error was considered negligible [3-5]. For all these reasons, bias in our AI (artificial intelligence) algorithms is highly unlikely. We, however, agree that admission to the ICU can be related to many factors. One strength of our study is that it analyzes the usual clinical practice in the whole population cared for in an entire health care region of Spain during a period when the lack of beds was not a limiting factor. It was not a sample—it was the entire population. Finally, our study objective was not mortality. In other studies, when

XSL•FO
RenderX

we addressed mortality, the study period was extended to reliably collect this variable [6,7].

Conflicts of Interest

None declared.

References

1. Martos Pérez F, Gomez Huelgas R, Martín Escalante MD, Casas Rojo JM. Minimizing Selection and Classification Biases. Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing". J Med Internet Res 2021 May;23(5):e27142. [doi: 10.2196/27142]
2. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. J Med Internet Res 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: 10.2196/21801] [Medline: 33090964]
3. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? Eur J Radiol 2020 May;126:108961. [doi: 10.1016/j.ejrad.2020.108961] [Medline: 32229322]
4. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
5. Xu J, Wu R, Huang H, Zheng W, Ren X, Wu N, et al. Computed Tomographic Imaging of 3 Patients With Coronavirus Disease 2019 Pneumonia With Negative Virus Real-time Reverse-Transcription Polymerase Chain Reaction Test. Clin Infect Dis 2020 Jul 28;71(15):850-852 [FREE Full text] [doi: 10.1093/cid/ciaa207] [Medline: 32232429]
6. Graziani D, Soriano JB, Del Rio-Bermudez C, Morena D, Díaz T, Castillo M, et al. Characteristics and Prognosis of COVID-19 in Patients with COPD. J Clin Med 2020 Oct 12;9(10):E3259 [FREE Full text] [doi: 10.3390/jcm9103259] [Medline: 33053774]
7. Izquierdo JL, Almonacid C, González Y, Del Rio-Bermudez C, Ancochea J, Cárdenas R, et al. The impact of COVID-19 on patients with asthma. Eur Respir J 2021 Mar;57(3):57 [FREE Full text] [doi: 10.1183/13993003.03142-2020] [Medline: 33154029]

Abbreviations

AI: artifical intelligence
ICU: intensive care unit
PCR: polymerase chain reaction

Edited by T Derrick; submitted 06.04.21; this is a non–peer-reviewed article;accepted 13.05.21; published 26.05.21.

Please cite as:
Izquierdo JL, Soriano JB
Authors' Reply to: Minimizing Selection and Classification Biases Comment on "Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing"
J Med Internet Res 2021;23(5):e29405
URL: https://www.jmir.org/2021/5/e29405
doi:10.2196/29405
PMID:33989164

©Jose Luis Izquierdo, Joan B Soriano. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2021. 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

Letter to the Editor

Objective Outcomes Evaluation of Innovative Digital Health Curricula. Comment on "Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study"

Alexander Grzeska[1*]; Shan Ali[1*], MD; Tomasz Szmuda[1], MD, PhD; Paweł Słoniewski[1], MD, PhD

Medical University of Gdansk, Gdańsk, Poland

[*]these authors contributed equally

Corresponding Author:
Alexander Grzeska
Medical University of Gdansk
ul M Skłodowskiej-Curie 3a
Gdańsk, 80-210
Poland
Phone: 48 572642516
Email: alex.grzeska@gumed.edu.pl

Related Article:

Comment on: https://www.jmir.org/2020/10/e22161/

(J Med Internet Res 2021;23(5):e26034) doi:10.2196/26034

KEYWORDS

digital health; eHealth; mHealth; digital health education; elective module; eHealth education; curriculum; medical school; digital health mindset; qualitative research; interview; survey

We read with great interest the study by Poncette et al [1] on the development of a curricular module for digital health and undergraduate medical competencies.

The paper [1] illuminates the importance of implementing digital health training for future physicians in the curricula of medical universities. While the paper discusses the benefits of building a critical and experience-based mindset among future physicians, we believe that measuring the performance of and skills acquired by students before and after participation may augment the development and implementation of digital health competencies in the medical curricula.

A survey found high levels of satisfaction among students participating in the elective module. Confidence levels, as well as the acquisition of knowledge and skills in digital health, were reported to be high. However, this analysis is subjective as it is based on the student's own opinion of their knowledge. In order to obtain more objective results, we recommend the assessment of individual skills by a standardized exam before and after participating in the module. Supplementing the opinion-based survey with exam performance may shed light on the actual perceived level of student competence while reducing selection bias. This would highlight the effectiveness of the course and give medical departments and universities more confidence to adopt digital health competencies in their curricula.

We commend Poncette et al [1] not only for their findings but for how they encourage a more interactive and nonclassical way of teaching. We hope that the high satisfaction among students may encourage medical universities to adopt this new model of teaching. The passive transfer of knowledge has been the dominating system used in medical universities [2] whereas active learning methods such as a "peer-to-peer" [1] teaching approach with an open discussion is rarely found, thus, resulting in high levels of satisfaction [3]. Therefore, we hope that this paper may persuade medical universities to adopt this new teaching approach.

The introduction of digital health competencies into the medical curricula remains key to the success of future physicians as shown by Poncette et al [1]. Therefore, we advocate for more objective student evaluation methods—before and after participating in learning modules—so that traditional medical institutions may be compelled to adopt this new teaching approach.

XSL•FO
RenderX

Editorial Notice

The corresponding author of "Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study" declined to respond to this letter.

Conflicts of Interest

None declared.

References

1. Poncette AS, Glauert DL, Mosch L, Braune K, Balzer F, Back DA. Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study. J Med Internet Res 2020 Oct 29;22(10):e22161 [FREE Full text] [doi: 10.2196/22161] [Medline: 33118935]
2. Graffam B. Active learning in medical education: strategies for beginning implementation. Med Teach 2007 Feb;29(1):38-42. [doi: 10.1080/01421590601176398] [Medline: 17538832]
3. Schmidt HG, Cohen-Schotanus J, Arends LR. Impact of problem-based, active learning on graduation rates for 10 generations of Dutch medical students. Med Educ 2009 Mar;43(3):211-218. [doi: 10.1111/j.1365-2923.2008.03287.x] [Medline: 19250347]

Edited by T Derrick; submitted 25.11.20; peer-reviewed by D Back, E Da Silva; comments to author 13.05.21; revised version received 15.05.21; accepted 17.05.21; published 28.05.21.

Please cite as:
Grzeska A, Ali S, Szmuda T, Słoniewski P
Objective Outcomes Evaluation of Innovative Digital Health Curricula. Comment on "Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study"
J Med Internet Res 2021;23(5):e26034
URL: https://www.jmir.org/2021/5/e26034
doi:10.2196/26034
PMID:34047706

©Alexander Grzeska, Shan Ali, Tomasz Szmuda, Paweł Słoniewski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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

Letter to the Editor

Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings"

Daniel Melo De Oliveira Campos[1*], MSc; Umberto Laino Fulco[1*], PhD; Jonas Ivan Nobre Oliveira[1*], PhD

Universidade Federal do Rio Grande do Norte, Natal, Brazil

[*]all authors contributed equally

Corresponding Author:
Jonas Ivan Nobre Oliveira, PhD
Universidade Federal do Rio Grande do Norte
Departamento de Biofísica e Farmacologia
Natal, 59072-970
Brazil
Phone: 55 8432153793
Email: jonasivan@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2020/11/e23449/

Comment in: https://www.jmir.org/2021/5/e29507/

(J Med Internet Res 2021;23(5):e28666) doi:10.2196/28666

KEYWORDS

coronavirus; COVID-19; pandemic; scientific publishing; PubMed; literature searching; research; literature; search; performance; search strategy

Recently, a very interesting study on the performance of different search strategies for COVID-19 records in PubMed was published in the *Journal of Medical Internet Research* [1]. In this article, Lazarus et al compared the performance of PubMed's one-click search option with both simpler and more complex search strings. Novice and expert searchers do well to keep these in mind when searching. For instance, a search strategy for a review is a time-consuming endeavor, and energy spent on locating relevant controlled vocabulary and keywords can be undermined by errors in formatting, compilation, and translation of these terms. Unfortunately, the presence of these errors is extremely common even among published studies. Sampson and McGowan [2] reviewed studies published in Cochrane and discovered that 90.5% of their sample had a search strategy that contained one or more errors. Some related to errors regarding term, or term variant, identifications, but others pertained to errors in the formatting and basic compilation of the terms. The latter category included Boolean errors (19%), incorrect line numbers (1.6%), the use of Medical Subject Headings (MeSH) and free text terms combined on the same line (20.6%), and the search strategy not being appropriately translated for other databases (20.6%).

In 2018, a study with a random sample of 70 Cochrane Reviews found problems in the design of the search strategies in 73% of reviews, and 53% of these contained problems that could limit both the sensitivity and precision of the search [3]. Recently, Salvador-Oliván et al (2019) [4] evaluated the search strategies of 137 systematic reviews in PubMed to identify errors, analyze their impact on information retrieval, and propose solutions. The results of this study reveal that the percentage of search strategies that contain various types of errors is quite high (92.7%) and that 78.1% of these errors affect recall. Although a substantial proportion of the errors came from inadequate identification of terms, errors were also introduced at the formatting level, with an absence of field tags (21.2%) and lack or incorrect use of quotation marks (5.8%), Boolean operators (1.5%), and parentheses (5.1%) [4].

As to be expected, some errors have graver effects on results than others. Errors that have no effect at all on the number of results include redundant terms and morphological repetition; these "search errors" do not affect recall or negatively affect information retrieval with respect to either recall or precision.

An example of redundancy is as follows: "2019 novel coronavirus disease"[tw] OR "2019 novel coronavirus infection"[tw] OR "2019-nCoV disease"[tw] OR "2019-nCoV infection"[tw] OR "COVID-19 pandemic"[tw] OR "COVID-19 virus disease"[tw] OR "COVID-19 virus infection"[tw] OR

XSL•FO
RenderX

"COVID19"[tw] OR "SARS-CoV-2 infection"[tw] OR "coronavirus disease 2019"[tw] OR "coronavirus disease-19"[tw] OR "COVID-19 pandemic"[tw] OR "COVID-19"[tw]. Authors justify redundancy because the decision to include or exclude terms depends on the references retrieved, as the effect of the terms on the results is impossible to predict. However, it is known beforehand that the first 11 terms in a PubMed search can be easily discarded because using the 12th variation will cover all 11, so other terms are unnecessary.

In terms of the search process, tools pertaining to data mining have been developed to help librarians identify relevant terms. Some text-mining approaches have been documented by Stansfield et al [5], including TFIDF, Termine, and BibExcel. Also recommended are librarian tools that often have a particular focus on the MeSH thesaurus, such as PubMed PubReMiner [6] and Yale MeSH Analyzer [7] for keywords and controlled vocabulary.

Created and updated by the United States National Library of Medicine, MeSH vocabulary is used by the ClinicalTrials.gov registry to classify which diseases are studied by the trials registered in its database. This hierarchically organized terminology for indexing and cataloging of biomedical information is divided into four types of terms. The main terms are the "headings" (also known as MeSH headings or descriptors), which describe the subject of each article. Most of these are accompanied by a list of synonyms or very similar terms (known as entry terms). When performing a MEDLINE search via PubMed, entry terms are automatically translated into (ie, mapped to) the corresponding descriptors with a good degree of reliability. In this sense, we highlighted the importance of using the controlled vocabulary "COVID-19" (unique id: C000657245) and "SARS-CoV-2" (unique id: D000086402) in PubMed searches focused on COVID-19–related studies, and not the set of terms (search 1, 2, 3, 6, 7, and 8) analyzed by Lazarus and collaborators [1].

Redundant terms in a search strategy do not affect the retrieval of information; however, the principle of parsimony instructs us to eliminate that which is unnecessary. Applied to information retrieval, this principle prompts us to eliminate any terms or phrases from a search strategy that do not retrieve or provide new records, as they are thus unnecessary.

Acknowledgments

This work was partially financed by the Brazilian Research Agencies CAPES (PNPD) and CNPq (Finance Code 001).

Conflicts of Interest

None declared.

References

1. Lazarus JV, Palayew A, Rasmussen LN, Andersen TH, Nicholson J, Norgaard O. Searching PubMed to retrieve publications on the COVID-19 pandemic: comparative analysis of search strings. J Med Internet Res 2020 Nov 26;22(11):e23449 [FREE Full text] [doi: 10.2196/23449] [Medline: 33197230]
2. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. J Clin Epidemiol 2006 Oct;59(10):1057-1063. [doi: 10.1016/j.jclinepi.2006.01.007] [Medline: 16980145]
3. Franco JVA, Garrote VL, Escobar Liquitay CM, Vietto V. Identification of problems in search strategies in Cochrane Reviews. Res Synth Methods 2018 Sep;9(3):408-416. [doi: 10.1002/jrsm.1302] [Medline: 29761662]
4. Salvador-Oliván J, Marco-Cuenca G, Arquero-Avilés R. Errors in search strategies used in systematic reviews and their effects on information retrieval. J Med Libr Assoc 2019 Apr 15;107(2):210. [doi: 10.5195/jmla.2019.567]
5. Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: a discussion of some methods and challenges. Res Synth Methods 2017 Sep;8(3):355-365. [doi: 10.1002/jrsm.1250] [Medline: 28660680]
6. Slater L. PubMed PubReMiner. J Can Health Libr Assoc 2014 Jul 21;33(2):106. [doi: 10.5596/c2012-014]
7. Hocking R. Yale MeSH Analyzer. J Can Health Libr Assoc 2017 Dec 01;38(3). [doi: 10.29173/jchla/jabsc.v38i3.29336]

Abbreviations

MeSH: Medical Subject Headings

XSL•FO
RenderX

Edited by T Derrick; submitted 10.03.21; this is a non–peer-reviewed article;accepted 13.05.21; published 28.05.21.

Please cite as:
Campos DMDO, Fulco UL, Oliveira JIN
Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings"
J Med Internet Res 2021;23(5):e28666
URL: https://www.jmir.org/2021/5/e28666
doi:10.2196/28666
PMID:33989165

©Daniel Melo De Oliveira Campos, Umberto Laino Fulco, Jonas Ivan Nobre Oliveira. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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

Letter to the Editor

Authors' Reply to: Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings"

Lauge Neimann Rasmussen[1]; Ole Norgaard[1]; Tue Helms Andersen[1]; Adam Palayew[2]; Joey Nicholson[3]; Jeffrey V Lazarus[4], PhD

[1]Danish Diabetes Knowledge Center, Steno Diabetes Center Copenhagen, Gentofte, Denmark

[2]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[3]NYU Langone Health, NYU Grossman School of Medicine, NYU Health Sciences Library, New York, NY, United States

[4]Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain

Corresponding Author:
Jeffrey V Lazarus, PhD
Barcelona Institute for Global Health (ISGlobal), Hospital Clínic
University of Barcelona
Calle del Rosellón 132
Barcelona, 08036
Spain
Email: Jeffrey.Lazarus@isglobal.org

Related Articles:

Comment on: https://www.jmir.org/2020/11/e28666/

Comment on: https://www.jmir.org/2020/11/e23449/

(J Med Internet Res 2021;23(5):e29507) doi:10.2196/29507

KEYWORDS

coronavirus; COVID-19; pandemic; scientific publishing; PubMed; literature searching; research; literature; search; performance; search strategy

We appreciate the interest in our analyses and dedication to informing clinicians, reviewers, information specialists, and others about searching for articles in PubMed.

The letter authors [1] highlight the importance of avoiding search errors and the principle of parsimony in formulating search strings, encouraging searchers "to eliminate any terms or phrases from a search strategy that do not retrieve or provide new records, as they are thus unnecessary." We support the ambition and principle of parsimony, agreeing that redundant terms should be avoided. In our view, search 1 contains no redundant terms [2]. It is unclear to us what search the letter authors refer to, and we encourage them to specify their critique regarding this. To avoid misunderstandings, the example search string in the letter does not originate from any that we tested.

Our analysis showed how term choice and combinations impacted the sensitivity, precision, and F-score of selected search strings in PubMed in the 10 weeks from when the World Health Organization declared COVID-19 a public health emergency of international concern. Our results demonstrated that the more elaborate searches 1 and 2 had higher sensitivity than the ones singled out in the letter (searches 4 and 5). Please note that searches 3 and 4 were in practice identical, which some readers may overlook. In both the abstract and article, we highlighted the value of applying the single-term search "COVID-19" (searches 3 and 4) for everyday searches. However, as reflected by the F-score, searches 1 and 2 performed best.

We acknowledge that the differences between some of the analyzed search strings are minor to some PubMed users and for some search purposes. For Cochrane-style systematic reviews, we recommend the more comprehensive search strings with the higher sensitivity and specificity unless resources are scarce. For everyday informational needs, less comprehensive searches may suffice.

Our study had limitations, importantly the timeframe represented. As COVID-19 has evolved (eg, variants have been detected and have spread), new terms may be pertinent to add to search strings, including relevant Medical Subject Headings (MeSH) terms, which were not available during our search. Instead, we used all available Supplementary Concepts identified

XSL•FO
RenderX

as relevant to COVID-19 in June 2020. Since the beginning of 2021, COVID-19–relevant MeSH terms have been available [3], which we would include if we were developing a comprehensive search string now.

We note that the suggestion from the letter authors to include the controlled vocabulary term "COVID-19" (unique ID: C000657245) would not activate the MeSH term "COVID-19" (unique ID: D000086382). Instead, it activates the obsolete Supplementary Concept "COVID-19", which we included in our search string.

The change in availability of relevant controlled vocabulary terms highlights the need to quickly add additional features, enabling PubMed users to perform effective searches during a pandemic. In line with our suggestion that the National Library of Medicine add a new subject filter, such as covid-19[sb], to address this issue, they have recently added several COVID-19 filters, such as LitCGeneral[filter], that may become useful [4]. For future pandemics, we hope that such features would be made available sooner and that diseases and their variants would be named earlier.

Conflicts of Interest

LNR, ON, and THA are employed at Steno Diabetes Center Copenhagen, a public hospital and research institution under the Capital Region of Denmark, which is partly funded by a grant from the Novo Nordisk Foundation. The funders had no role in this work. There are no other conflicts to declare.

References

1. Campos DMDO, Fulco UL, Oliveira JIN. Redundancy of terms in search strategies to identify COVID-19–related articles in PubMed: comment on "Searching pubmed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings". J Med Internet Res 2021 May 28;23(5):e28666 [FREE Full text] [doi: 10.2196/28666] [Medline: 33989165]
2. Lazarus JV, Palayew A, Rasmussen LN, Andersen TH, Nicholson J, Norgaard O. Searching PubMed to retrieve publications on the COVID-19 pandemic: comparative analysis of search strings. J Med Internet Res 2020 Nov 26;22(11):e23449 [FREE Full text] [doi: 10.2196/23449] [Medline: 33197230]
3. Tybaert S. MEDLINE data changes—2021. NLM Tech Bull 2020;437:e3a [FREE Full text]
4. PubMed user guide - COVID-19 article filters. National Library of Medicine. URL: https://pubmed.ncbi.nlm.nih.gov/help/#covid19-article-filters [accessed 2020-04-08]

Edited by T Derrick; submitted 09.04.21; this is a non–peer-reviewed article;accepted 13.05.21; published 28.05.21.

Please cite as:
Rasmussen LN, Norgaard O, Andersen TH, Palayew A, Nicholson J, Lazarus JV
Authors' Reply to: Redundancy of Terms in Search Strategies. Comment on "Searching PubMed to Retrieve Publications on the COVID-19 Pandemic: Comparative Analysis of Search Strings"
J Med Internet Res 2021;23(5):e29507
URL: https://www.jmir.org/2021/5/e29507
doi:10.2196/29507
PMID:33989168

©Lauge Neimann Rasmussen, Ole Norgaard, Tue Helms Andersen, Adam Palayew, Joey Nicholson, Jeffrey V Lazarus. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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: Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study

Stanley Xu[1], PhD; Sungching Glenn[1], MSc; Lina Sy[1], MPH; Lei Qian[1], PhD; Vennis Hong[1], MPH; Denison S Ryan[1], MPH; Steven Jacobsen[1], MD, PhD

Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

Corresponding Author:
Stanley Xu, PhD
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S Los Robles Ave, 5th Floor
Pasadena, CA, 91101
United States
Phone: 1 6263726807
Fax: 1 626 564 3409
Email: stan.xu@kp.org

Related Article:

Correction of: https://www.jmir.org/2021/4/e26558/

(J Med Internet Res 2021;23(5):e30101) doi:10.2196/30101

In "Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study" (J Med Internet Res 2021;23(4):e26558) the authors noted two errors.

In the originally published manuscript, the following sentence appeared under the "Data Source and Identification of Visits" section:

> *They were not considered telehealth consultations in this study.*

This sentence has been corrected to:

> *They were not considered telehealth visits in this study.*

In the originally published manuscript, the following sentence was included under the "Principal Findings" section:

> *We observed that KPSC membership generally remained stable during the pandemic, largely owing to the KPSC's decision to not cancel health coverage for groups or individuals who could not pay for most of the study period [16].*

In the corrected version, the citation to Reference 16 has been removed from this sentence.

The correction will appear in the online version of the paper on the JMIR Publications website on May 5, 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 30.04.21; this is a non–peer-reviewed article;accepted 30.04.21; published 05.05.21.

Please cite as:
Xu S, Glenn S, Sy L, Qian L, Hong V, Ryan DS, Jacobsen S
Correction: Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study
J Med Internet Res 2021;23(5):e30101
URL: https://www.jmir.org/2021/5/e30101
doi:10.2196/30101
PMID:33950848

XSL•FO
RenderX

©Stanley Xu, Sungching Glenn, Lina Sy, Lei Qian, Vennis Hong, Denison S Ryan, Steven Jacobsen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2021. 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: Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study

Malte L Schmieding[1,2], MD; Rudolf Mörgeli[1], MD; Maike A L Schmieding[3]; Markus A Feufel[4*], Dipl-Ing (FH), MSc, PhD; Felix Balzer[1,2*], MSc, PhD, MD

[1]Department of Anesthesiology and Operative Intensive Care, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[3]Department of Biology, Chemistry, and Pharmacy, Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany

[4]Department of Psychology and Ergonomics (IPA), Division of Ergonomics, Technische Universität Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Felix Balzer, MSc, PhD, MD
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 570425
Email: felix.balzer@charite.de

Related Article:

Correction of: https://www.jmir.org/2021/3/e24475/

(J Med Internet Res 2021;23(5):e30215) doi:10.2196/30215

In “Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study” (J Med Internet Res 2021;23(3):e24475) the authors noted two errors.

In the originally published manuscript, Reference 24 in the reference list was cited incorrectly. This citation has now been corrected to the reference listed below [1].

In the originally published paper, there was an error in the phone number of the Corresponding Author. The number has been corrected from "49 30 450 5704" to "49 30 450 570425".

The correction will appear in the online version of the paper on the JMIR Publications website on May 6, 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.

Reference

1. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Data set supplementing "Benchmarking triage capability of symptom checkers against that of medical laypersons: Survey study". Zenodo. 2021. URL: https://zenodo.org/record/4454538#.YJL3EqlKj0p [accessed 2021-05-05]

XSL•FO
RenderX

Submitted 05.05.21; this is a non–peer-reviewed article;accepted 05.05.21; published 06.05.21.

Please cite as:
Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F
Correction: Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study
J Med Internet Res 2021;23(5):e30215
URL: https://www.jmir.org/2021/5/e30215
doi:10.2196/30215
PMID:33956638

©Malte L Schmieding, Rudolf Mörgeli, Maike A L Schmieding, Markus A Feufel, Felix Balzer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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: Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey

Fengyun Yu[1,2*], BSc; Pascal Geldsetzer[2,3*], MD, ScD; Anne Meierkord[4], MD; Juntao Yang[5], PhD; Qiushi Chen[6], PhD; Lirui Jiao[7], BA; Nadeem E Abou-Arraj[8], MD; An Pan[9], PhD; Chen Wang[10,11,12*], MD, PhD; Till Bärnighausen[2,10*], MD, ScD; Simiao Chen[2,10*], ScD

[1]Department of Industrial Engineering, Tsinghua University, Beijing, China

[2]Heidelberg Institute of Global Health, Faculty of Medicine and University Hospital, Heidelberg University, Heidelberg, Germany

[3]Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[4]Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[5]State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[6]The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA, United States

[7]Reed College, Portland, OR, United States

[8]Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[9]School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[10]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[11]National Clinical Research Center for Respiratory Diseases, Beijing, China

[12]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China–Japan Friendship Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Simiao Chen, ScD
Heidelberg Institute of Global Health
Faculty of Medicine and University Hospital
Heidelberg University
Im Neuenheimer Feld 130.3
Heidelberg
Germany
Phone: 49 15207926273
Email: simiao.chen@uni-heidelberg.de

Related Article:

Correction of: https://www.jmir.org/2021/4/e26940/

(J Med Internet Res 2021;23(5):e30100) doi:10.2196/30100

In “Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey” (J Med Internet Res 2021;23(4):e26940) the authors noted three errors.

In the originally published manuscript, authors Chen Wang, Till Bärnighausen, and Simiao Chen were noted as having contributed equally to the manuscript.

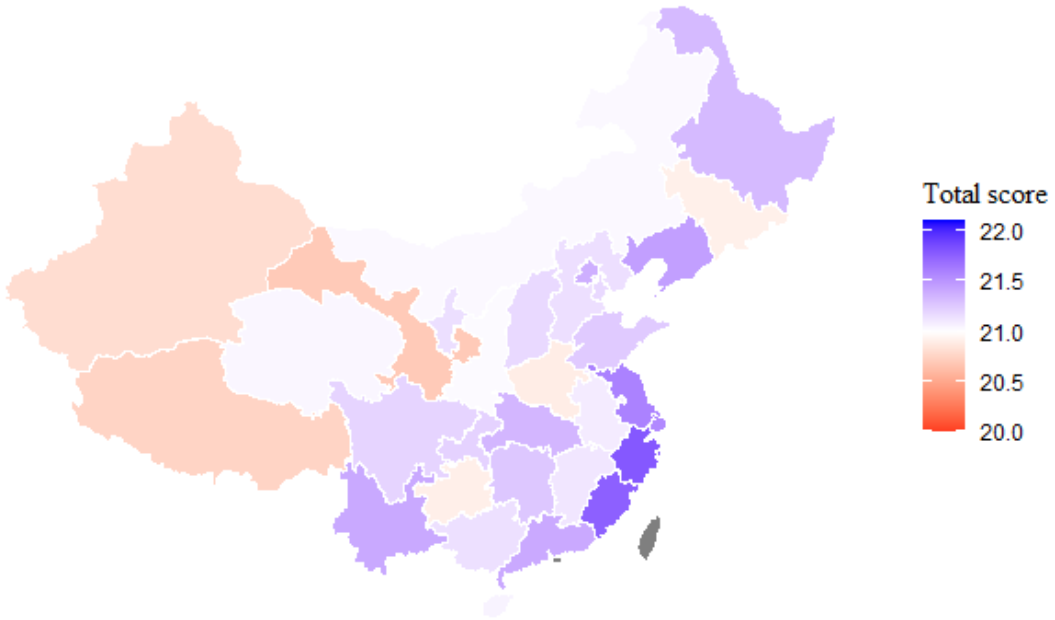
This has been corrected to note that authors Fengyun Yu, Pascal Geldsetzer, Chen Wang, Till Bärnighausen, and Simiao Chen all contributed equally to the manuscript.

Additionally, the following figures have been revised with minor changes in scale and legend title.

- Figure 1: Map showing the mean overall knowledge score by province.
- Supplementary Figure A3: Map showing the mean overall knowledge score by province when excluding participants who reported looking up an answer online (Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Map showing the mean overall knowledge score by province.



The correction will appear in the online version of the paper on the JMIR Publications website on May 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.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 121 KB - jmir_v23i5e30100_app1.docx]

Submitted 30.04.21; this is a non–peer-reviewed article;accepted 30.04.21; published 12.05.21.

Please cite as:
Yu F, Geldsetzer P, Meierkord A, Yang J, Chen Q, Jiao L, Abou-Arraj NE, Pan A, Wang C, Bärnighausen T, Chen S
Correction: Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey
J Med Internet Res 2021;23(5):e30100
URL: https://www.jmir.org/2021/5/e30100
doi:10.2196/30100
PMID:33979299

©Fengyun Yu, Pascal Geldsetzer, Anne Meierkord, Juntao Yang, Qiushi Chen, Lirui Jiao, Nadeem E Abou-Arraj, An Pan, Chen Wang, Till Bärnighausen, Simiao Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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

Analysis of COVID-19 Transmission Sources in France by Self-Assessment Before and After the Partial Lockdown: Observational Study

Fabrice Denis[1], MD, PhD; Anne-Lise Septans[2], PhD; Florian Le Goff[3], MSc; Stephan Jeanneau[4], MSc; François-Xavier Lescure[5], MD, PhD

[1]Inter-regional Cancer Institut Jean Bernard - ELSAN, Le Mans, France

[2]Weprom, Angers, France

[3]Kelindi, Lille, France

[4]Adobis Group, Grenoble, France

[5]Infectious and Tropical Diseases Department, Bichat-Claude Bernard University Hospital, Assistance Publique – Hôpitaux de Paris, Paris, France

Corresponding Author:
Fabrice Denis, MD, PhD
Inter-regional Cancer Institut Jean Bernard - ELSAN
9 Rue Beauverger
Le Mans,
France
Phone: 33 0684190480
Email: fabrice.denis.home@gmail.com

Abstract

Background: We developed a questionnaire on a web application for analyzing COVID-19 contamination circumstances in France during the second wave of the pandemic.

Objective: This study aims to analyze the impact on contamination characteristics before and after the second partial lockdown in France to adapt public health restrictions to further prevent pandemic surges.

Methods: Between December 15 and 24, 2020, after a national media campaign, users of the sourcecovid.fr web application were asked questions about their own or a close relative's COVID-19 contamination after August 15, 2020, in France. The data of the contamination's circumstances were assessed and compared before and after the second partial lockdown, which occurred on October 25, 2020, during the second wave of the pandemic and was ongoing on December 24, 2020.

Results: As of December 24, 2020, 441,000 connections on the web application were observed. A total of 2218 questionnaires were assessable for analysis. About 61.8% (n=1309) of the participants were sure of their contamination origin, and 38.2% (n=809) thought they knew it. The median age of users was 43.0 (IQR 32-56) years, and 50.7% (n=1073) were male. The median incubation time of the assessed cohort was 4.0 (IQR 3-5) days. Private areas (family's or friend's house) were the main source of contamination (1048/2090, 50.2%), followed by work colleagues (579/2090, 27.7%). The main time of day for the contamination was the evening (339/961, 35.3%) before the lockdown and was reduced to 18.2% (86/473) after the lockdown (P<.001). The person who transmitted the virus to the user before and after the lockdown was significantly different (P<.001): a friend (382/1317, 29% vs 109/773, 14.1%), a close relative (304/1317, 23.1% vs 253/773, 32.7%), or a work colleague (315/1317, 23.9% vs 264/773, 34.2%). The main location where the virus was transmitted to the users before and after the lockdown was significantly different too (P<.001): home (278/1305, 21.3% vs 194/760, 25.5%), work (293/1305, 22.5% vs 225/760, 29.6%), collective places (430/1305, 33% vs 114/760, 15%), and care centers (58/1305, 4.4% vs 74/760, 9.7%).

Conclusions: Modalities of transmissions significantly changed before and after the second lockdown in France. The main sources of contamination remained the private areas and with work colleagues. Work became the main location of contamination after the lockdown, whereas contaminations in collective places were strongly reduced.

Trial Registration: ClinicalTrials.gov NCT04670003; https://clinicaltrials.gov/ct2/show/NCT04670003

(J Med Internet Res 2021;23(5):e26932) doi:10.2196/26932

XSL•FO
RenderX

KEYWORDS

COVID-19; web application; digital health; analysis; transmission; France; self-assessment; lockdown; observational; survey; impact; public health

Introduction

Patient-reported outcome applications have been shown to improve the health outcomes of patients including decreasing mortality [1-3].

We developed and launched a self-assessment and participatory surveillance web application called maladiecoronavirus.fr during the growing phase of the COVID-19 pandemic in March 2020 in France. This self-triage tool aimed to help patients who were symptomatic to be directed toward the emergency call or the general practitioner after analysis of symptoms and comorbidities. We showed that data from this web application could be a relevant tool to reduce the burden on emergency call centers [4]. It also proved to be useful in monitoring COVID-19 spread during the whole pandemic, with time and spatial correlations between number of hospitalizations and daily reported anosmia by users being higher than large-scale reverse transcription polymerase chain reaction (RT-PCR) positive tests [5,6]. A national partial lockdown was initiated in France on October 25, 2020, against the second wave of the COVID-19 pandemic. Contrary to the first complete lockdown from March to May 2020, this one maintained scholar, professional, and shopping activities. However, the circumstances of virus transmissions before and after this lockdown are not well known in France. We thus developed a specific questionnaire on a web application (sourcecovid.fr) for COVID-19 source of contamination analysis in France in December 2020 just after the second wave of the pandemic [7]. The objective of this national survey was to analyze the impact on contamination circumstances before and after the second partial lockdown in France initiated on October 25, 2020, associated with the second wave of the pandemic to optimize health public policy to further pandemic surges.

Methods

Users of sourcecovid.fr were recruited via a national media campaign in France from December 15 to 25, 2020, including through social media, radio, and magazine campaigns, between December 15-18, 2020. Participants were recruited through the website. Respondents provided information on sociodemographic data, zip code, coexisting disorders anonymously, and the severity of their disease. Only symptomatic users were recruited. They were asked to enter data about their own contamination or the contamination of a close relative, about their sureness about the contamination's circumstances ("I am sure," "I think I know," or "I don't know"), and they also had to answer when, by who, and where they thought they (or the close relative) were contaminated. Users who answered "I don't know" were excluded from analysis.

Questionnaires were excluded from the analysis if completion duration was considered inconsistent (below 100 or above 800 seconds); if users were asymptomatic; if they did not know about the contamination's circumstances; and if contamination occurred before August 15, 2020, to reduce memory bias.

We excluded incubation times greater than 14 days from the analysis. Incubation time was calculated by comparing the date of presumed contamination and the date of first symptoms. The study was approved by the French National Health-Data Institute, which reviews ethical conduct of human participant research, data confidentiality, and safety. The website was not considered a medical device by regulatory authorities since no tracking was performed and data were anonymous. The web application did not have access to testing results. Access to the web application did not require a log-in or creating an account. The web application did not identify participants who responded several times.

Data of the contamination circumstances were assessed and compared before and after the second partial lockdown, which occurred on October 25, 2020, during the second wave of the pandemic and was ongoing on December 24, 2020. Fisher exact test was performed to assess changes in circumstances of contaminations.

Results

As of December 24, 2020, 441,000 connections on the web application were observed. There were 2118 questionnaires assessable for analysis; 61.8% (n=1309) of the users were sure of their contamination circumstance, and 38.2% (n=809) thought they knew it. Sureness was not different according to age (P=.43). The median age of users was 43.0 (IQR 32-56) years, and 48.3% (1073/2218) were female. The total population older than 65 years made up 12.5% (n=265) of users, and 4.4% (n=93) of the questionnaires concerned people younger than 18 years. The median incubation time was assessable in 1676 questionnaires and was 4.0 (IQR 3-5) days. Whatever the sureness, time incubation was not different (P=.36). Among the incubation sample, 41.7% (699/1676) declared a positive RT-PCR or antigenic test.

Mild or moderate infection was reported by 85.1% (n=1802) of the 2118 questionnaires, severe infection in 10.8% (n=229), and hospitalization in 4.6% (n=98).

The partial lockdown occurred on August 25, 2020, and was associated with an 80% reduction of daily contaminations (Figure 1).

During the period between August 15 and December 24, 2020, the private area (family and friends) was the main source of contamination in the 2090 questionnaires (n=1048, 50.2%) followed by work colleagues (n=579, 27.7%), or an unknown person (n=299, 14.3%), and 3.9% (n=83) did not know who contaminated them.

The lockdown occurred on October 25, 2020; 1334 questionnaires described the contamination's circumstances

XSL•FO
RenderX

between August 15 and October 24, 2020, and 784 described it between October 25 and December 24, 2020.

The person who transmitted the virus to the user before and after the lockdown was significantly different (P<.001): a friend (382/1317, 29% vs 109/773, 14.1%), a close relative (304/1317, 23.1% vs 253/773, 32.7%), and a work colleague (315/1317, 23.9% vs 264/773, 34.2%; P<.001; Figure 2).

Figure 1. Impact of the partial lockdown on daily reverse transcription polymerase chain reaction positive tests. Red area: national partial lockdown period initiated on October 25, 2020.

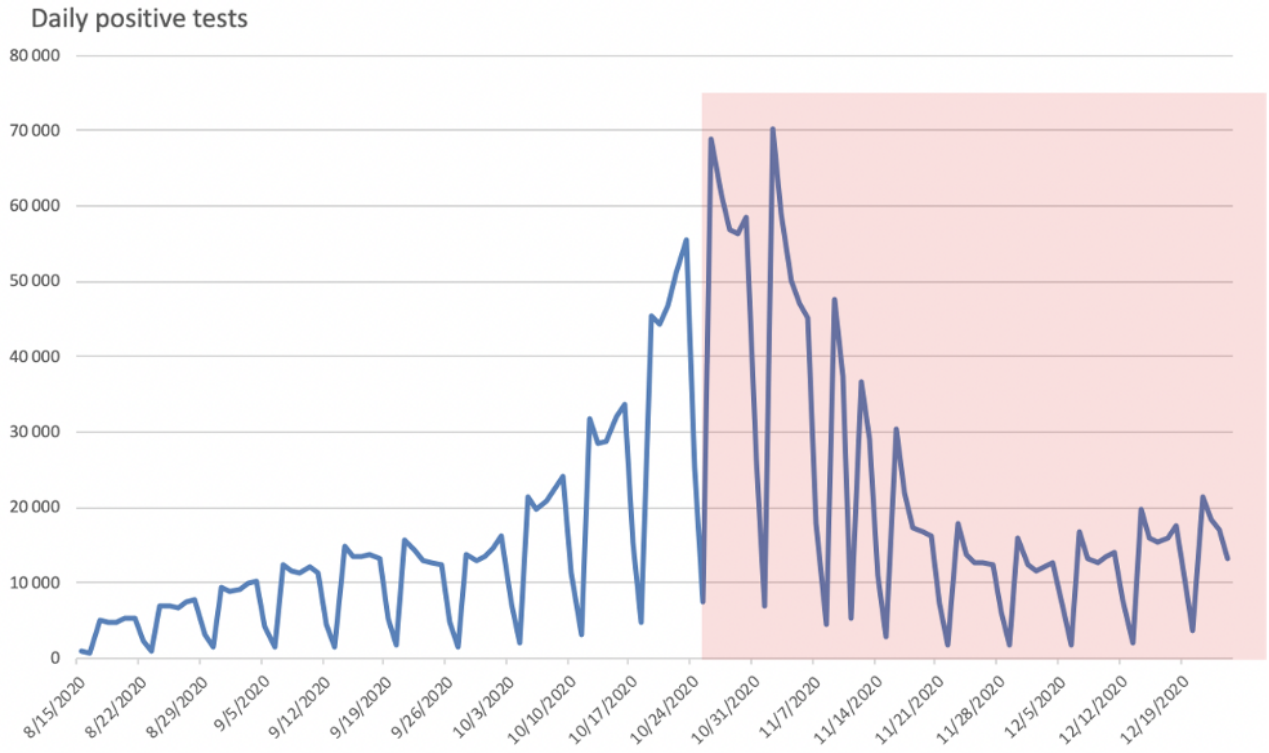
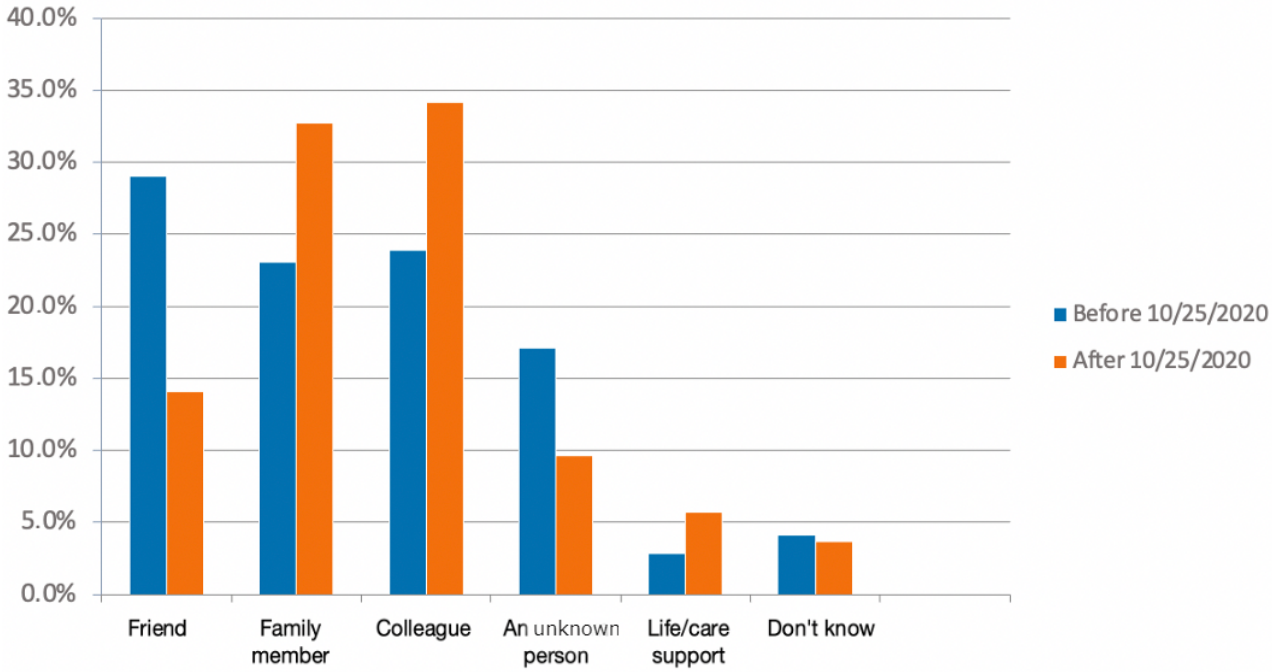


Figure 2. Contamination circumstances before versus after the partial lockdown, which was triggered in France on October 25, 2020. Answer to the question "Who contaminated you or your close relative?" (P<.001).



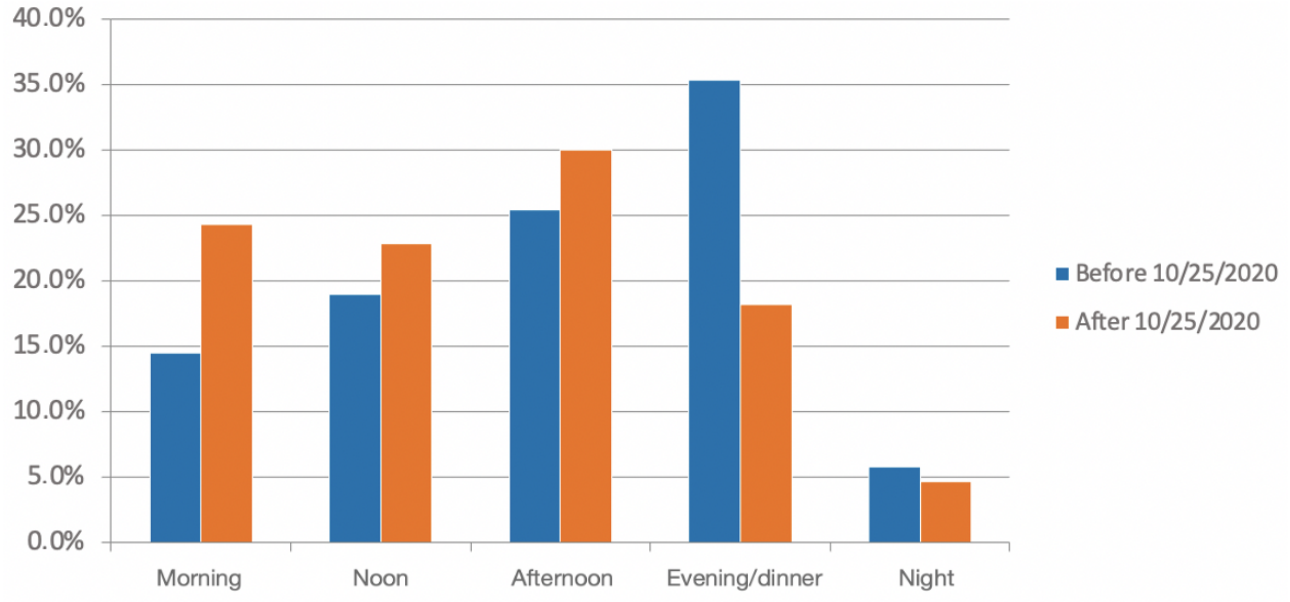
The distribution of responses regarding the people who had contaminated the user was different between users who were sure and users who thought they knew the origin of their contamination (P<.001). If professional relation was privileged for both group, users were more, in proportion, likely to think

XSL•FO
RenderX

it was an unknown person who infected them (151/790, 19.1% vs 148/1300, 11.4%).

The main time of contamination also changed after the lockdown. Among people who knew the time of contamination (n=1434; n=961 before the lockdown; n=473 after the lockdown), the main time was the evening (339/961, 35.3%) before the lockdown, which was reduced to 18.2% (86/473) after the lockdown (P<.001). The main time of contamination became the morning after the lockdown initiation. Morning and noon together became 47.2% (223/473) of the contamination times (Figure 3).

Figure 3. Contamination circumstances before and after the partial lockdown, which was triggered in France on October 25, 2020. Answer to the question "When do you think the contamination occurred?" (P<.001).



The distribution of responses regarding the moment of contamination was different according to the level of certainty expressed by the user (P=.005). Even if the preferred time of day was the evening, it was observed that users who were less sure estimated more than users who were sure to have contracted the virus in the morning: 21.5% (114/529) vs 15.5% (140/905), respectively.

The distribution of responses regarding the location of suspected contamination was different according to the sureness of contamination (P=.003). For users who though they knew where or when they were contaminated, they privileged the collective place compared to users who were sure: 29.8% (231/775) vs 24.3% (313/1290), respectively.

The location of suspected contamination by users of the web application questionnaire was home (own home, family's home, or friend's home) in 39.1% (510/1305) of the declarations before the lockdown and remained high after the lockdown: 43.3% (329/760). The other locations where the virus was transmitted to the users changed before versus after the lockdown (P<.001). It was increased in the work area (293/1305, 22.5% before the lockdown vs 225/760, 29.6% after it), reduced in collective places (430/1305, 33.0% vs 114/760, 15.0%), and increased in care centers (58/1305, 4.4% vs 74/760, 9.7%). Work became the main location of contamination after the lockdown (Figure 4).

Collective places where transmission occurred were significantly different before and after the lockdown (P<.001). The main collective places in terms of virus transmission before the lockdown were restaurants with a reduction from 27.0% (114/422) to 16.7% (19/114), bars (from 68/422, 16.1% to 6/114, 5.2%), parties (from 76/422, 18.0% to 4/114, 3.5%), and sports (from 42/422, 10.0% to 5/114, 4.4%). Among collective places, transportation and schools became the main collective places of transmissions after the partial lockdown: 17.5% (20/114) and 29.8% (34/114), respectively (Figure 5).

XSL•FO
RenderX

Figure 4. Contamination circumstances before and after the partial lockdown, which was triggered in France on October 25, 2020. Answer to the question "Where do you think the contamination occurred?" (*P*<.001).

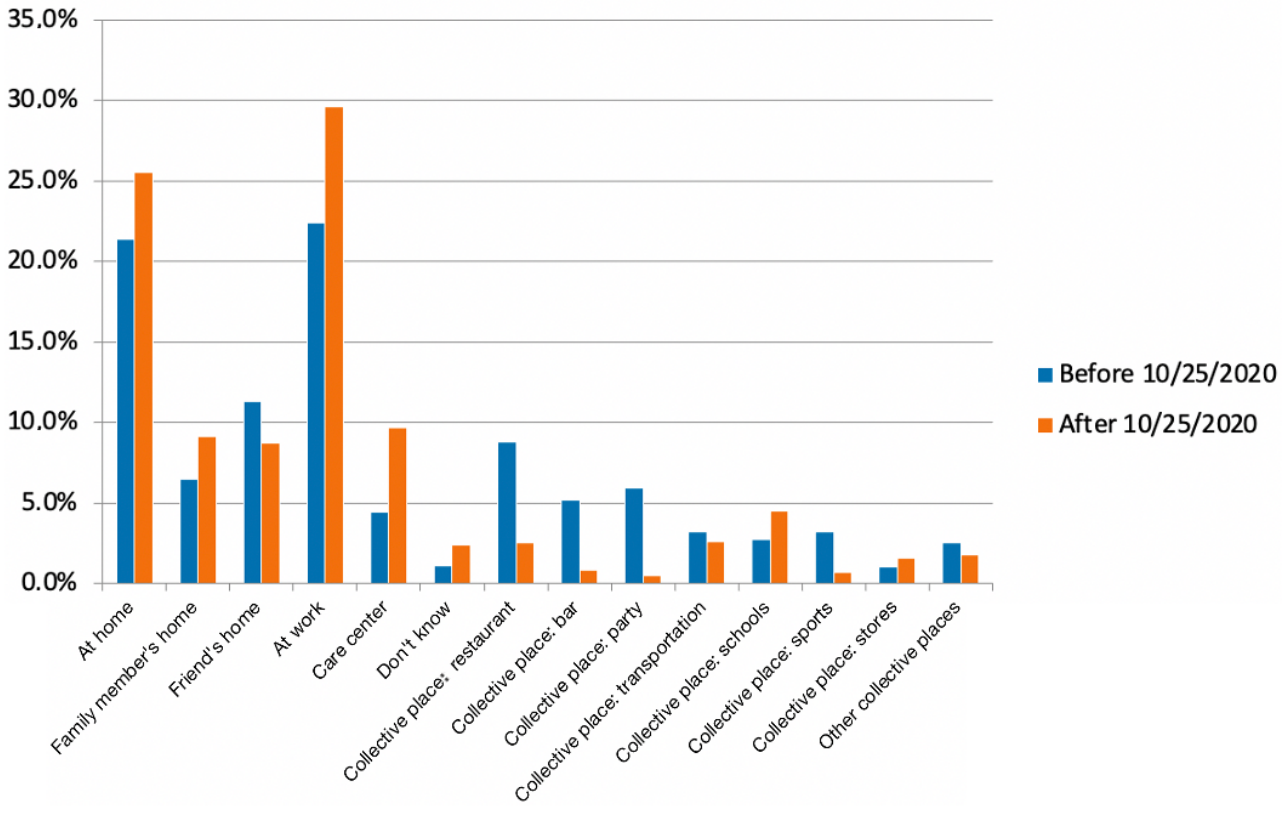
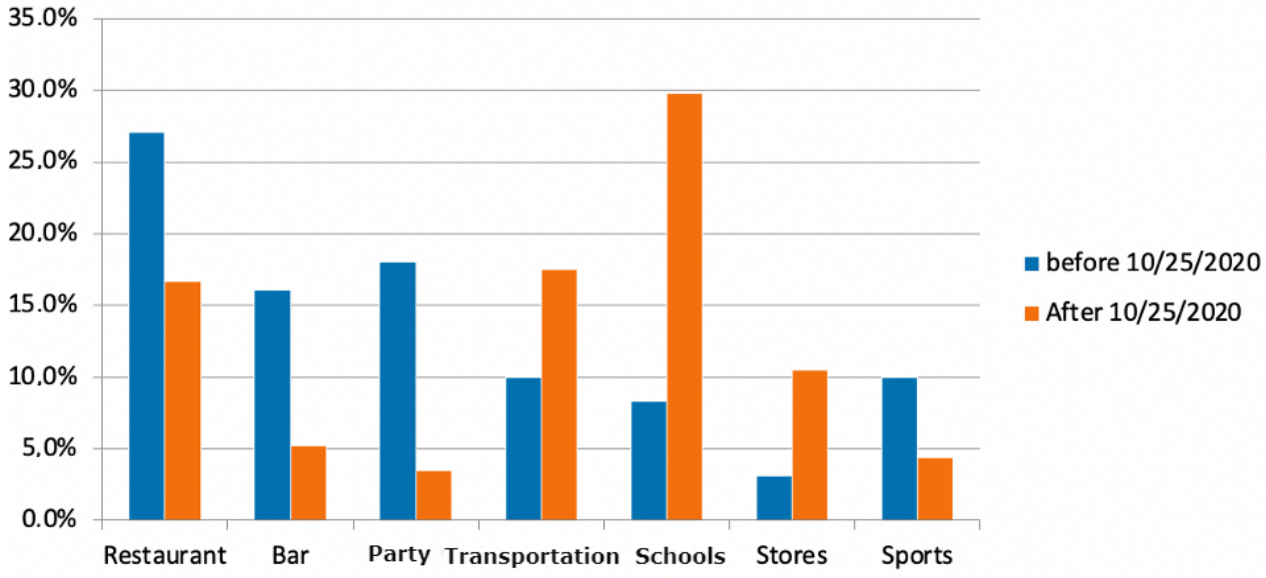


Figure 5. Main collective places concerned by contaminations before and after lockdown (*P*<.001); only principal location of collective area).



Among collective places, other collective places than those previously cited were reported by 4.5% (19/422) and 11.4% (13/114) of the users before and after the lockdown initiations, respectively (19/1334, 1.5% and 13/784, 1.7% of all contaminations, respectively).

For the total contaminations, restaurants were the source of only 2.5% (19/760) of infections after the lockdown versus 8.8% (114/1297) before the lockdown; bars were the source for 0.8% (6/760) of infections versus 5.2% (68/1297) before the lockdown; schools and stores were the source of 4.5% (34/760 vs 35/1297, 2.7%) and 1.6% (12/760 vs 13/1297, 1.0%, respectively) before the lockdown.

Discussion

Our results suggest that virus transmission occurred mainly in private areas (1048/2090, 50.2%) between August 15 and December 24, 2020, in France or by a colleague at work (579/2090, 27.7%). The partial lockdown, which occurred on

XSL•FO
RenderX

October 25, 2020, during the second wave of COVID-19, changed the circumstances of contamination.

Contamination by a friend was significantly reduced (382/773, 29.0%) before and (109/773, 14.1%) after the lockdown but increased from a family member or a colleague. Among people who knew the time of contamination, the main time of the day was the evening before the lockdown (339/961, 35.3%), and the morning and noon became the main time of contamination after the lockdown (86/473, 18.2%). The main location of suspected contamination by users of the web application questionnaire was private home (own home, family's home, or friend's home) in 39.1% (510/1305) of declarations before the lockdown and remained high after the lockdown (329/760, 43.3%).

Contaminations increased in the work area (293/1305, 22.5% vs 225/790, 29.6%; *P*<.001), reduced in collective places (430/1305, 33% vs 114/760, 15.0%; *P*<.001), and increased in care centers (58/1305, 4.4% vs 74/760, 9.7%; *P*<.001). Work became the main location of contamination after the lockdown. The main collective places impacted by the shutdown were restaurants, bars, and parties in which reduction of contaminations was significant.

Our results showed that contaminations occurred in collective places 33.0% (430/1305) of the time, and that the lockdown reduced it to 15.0% (114/760) of contaminations (ie, a 54.5% reduction), whereas daily anosmia reported on the national website maladiecoronavirus.fr showed an 80% reduction after the partial lockdown initiation. This can be explained by the voluntary decisions of people to reduce social meetings at a higher rate than government-imposed restrictions on activity during the lockdown and especially among friends, from which contaminations decreased from 29.0% to 14.1% [8]. As work and school were maintained during the shutdown, higher rates of contaminations were observed in those places. The partial lockdown concerned all cultural locations such as theaters or cinemas. We did not ask users of the web application sourcecovid.fr specifically if they thought they were contaminated in those sites but only "other locations." However, "other locations" was answered by nearly 1.5% of users before and after the lockdown, suggesting that cultural sites were not a significant source of contamination.

Although this study is based on unverifiable data, the quality of the data is consolidated by the incubation period calculated at 4.0 days, which is the median time reported in the literature; the equity between males and females; 10.8% (229/2118) of users with severe disease (15% in Guan et al [9]); the median age of users (43 years in our study, 47 years in Guan et al [9]); and 61.8% (1309/2118) of users reporting sureness of contamination circumstances. Moreover, results were consistent with published reports on excess of contamination risk in restaurants (times 2.4) or bars (times 3.9) and households (times 10), with a higher number of questionnaires in our study (n=2218) than in Fisher et al [10-12].

Our study has limitations. A user of the web application could use it several times and could have filled out more than one questionnaire. Memory bias could occur to users contaminated at the beginning of the study period, which explains the relative high number of users during the recent lockdown period.

The population older than 65 years made up 12.5% (265/2118) of users, whereas contamination circumstances were not the same as in the active population. Moreover, only 4.4% (93/2118) of questionnaires concerned people younger than 18 years. This has probably underestimated contaminations at school. We do not have enough data to assess geographic variations of the lockdown effect, especially to differentiate urban and rural area impact.

However, our study is the first to assess a partial lockdown effect on contamination circumstances and may help health authorities to adapt a policy of preventing COVID-19 spread.

Acknowledgments

We thank users for their participation in this study and Magali Balavoine, MSc (Weprom, Angers, France).

We also thank the Caisse Primaire d'Assurance Maladie de Sarthe, Le Mans Metropole (Stéphane Le Foll), Docaposte, Adobis group, and Le Figaro.fr.

Weprom designed and conducted the study; collected, managed, analyzed, and interpreted the data; prepared, reviewed, or approved the manuscript; and decided to submit the manuscript for publication.

Authors' Contributions

FD had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. DP also contributed toward the concept and design of the study. All authors were involved with acquisition, analysis, or interpretation of the data; drafting of the manuscript; and critical revision of the manuscript for important intellectual content. FD and ALS contributed toward the statistical analysis. FD, Weprom, Docaposte, and Kelindi were involved with administrative, technical, or material support. FD supervised the study.

Conflicts of Interest

All authors have completed and submitted the International Committee of Medical Journal Editors Form for Disclosure of Potential Conflicts of Interest. FD reports receiving personal fees from Astrazeneca, Ipsen, Kelindi, Pfizer, Chugai, and Roche. He is a cofounder of Kelindi. FLG is a cofounder of Kelindi. SJ is the founder of Adobis Group.

XSL•FO
RenderX

References

1. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
2. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
3. Basch E. Patient-reported outcomes - harnessing patients' voices to improve clinical care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
4. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure F, et al. Implementation of a self-triage web application for suspected COVID-19 and its impact on emergency call centers: observational study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]
5. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological observations on the association between anosmia and COVID-19 infection: analysis of data From a self-assessment web application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]
6. Denis F, Fontanet A, Le Douarin YM, Le Goff F, Jeanneau S, Lescure FX. A self-assessment web-based app to assess trends of the COVID-19 pandemic in France: observational study. J Med Internet Res 2021 Mar 12;23(3):e26182 [FREE Full text] [doi: 10.2196/26182] [Medline: 33709945]
7. SourceCovid.fr. URL: https://sourcecovid.fr [accessed 2020-12-15]
8. Goolsbee AC, Syverson C. Fear, lockdown, and diversion: comparing drivers of pandemic economic decline 2020. J Public Econ 2021 Jan;193:104311 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104311] [Medline: 33262548]
9. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/nejmoa2002032]
10. Fisher KA, Tenforde MW, Feldstein LR, Lindsell CJ, Shapiro NI, Files DC, CDC COVID-19 Response Team. Community and close contact exposures associated with COVID-19 among symptomatic adults ≥18 years in 11 outpatient health care facilities - United States, July 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1258-1264. [doi: 10.15585/mmwr.mm6936a5] [Medline: 32915165]
11. Lu J, Gu J, Li K, Xu C, Su W, Lai Z, et al. COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. Emerg Infect Dis 2020 Jul;26(7):1628-1631. [doi: 10.3201/eid2607.200764] [Medline: 32240078]
12. Lei H, Xu X, Xiao S, Wu X, Shu Y. Household transmission of COVID-19-a systematic review and meta-analysis. J Infect 2020 Dec;81(6):979-997 [FREE Full text] [doi: 10.1016/j.jinf.2020.08.033] [Medline: 32858069]

Abbreviations

RT-PCR: reverse transcription polymerase chain reaction

Edited by C Basch; submitted 07.01.21; peer-reviewed by E Stavaux, K Eason; comments to author 21.02.21; revised version received 22.02.21; accepted 16.04.21; published 04.05.21.

Please cite as:
Denis F, Septans AL, Le Goff F, Jeanneau S, Lescure FX
Analysis of COVID-19 Transmission Sources in France by Self-Assessment Before and After the Partial Lockdown: Observational Study
J Med Internet Res 2021;23(5):e26932
URL: https://www.jmir.org/2021/5/e26932
doi:10.2196/26932
PMID:33878018

©Fabrice Denis, Anne-Lise Septans, Florian Le Goff, Stephan Jeanneau, François-Xavier Lescure. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2021. 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 Secure, Continually Updating, Web Source Processing Pipeline to Support the Real-Time Data Synthesis and Analysis of Scientific Literature: Development and Validation Study

Uddhav Vaghela[1*], MBBS, BSc; Simon Rabinowicz[1*], MBBS, BSc; Paris Bratsos[1], MBBS, BSc; Guy Martin[1], PhD; Epameinondas Fritzilas[2], PhD; Sheraz Markar[1], PhD; Sanjay Purkayastha[1], PhD; Karl Stringer[3]; Harshdeep Singh[4], MSc; Charlie Llewellyn[2]; Debabrata Dutta[4], PhD; Jonathan M Clarke[1], PhD; Matthew Howard[2], PhD; PanSurg REDASA Curators[1,5*]; Ovidiu Serban[6], PhD; James Kinross[1], FRCS, PhD

[1]PanSurg Collaborative, Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Amazon Web Services UK Limited, London, United Kingdom

[3]MirrorWeb Limited, Manchester, United Kingdom

[4]Cloudwick Technologies, Newark, CA, United States

[5]see Acknowledgments, London, United Kingdom

[6]Data Science Institute, Imperial College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Ovidiu Serban, PhD
Data Science Institute
Imperial College London
William Penney Laboratory
South Kensington Campus
London
United Kingdom
Email: o.serban@imperial.ac.uk

Abstract

Background: The scale and quality of the global scientific response to the COVID-19 pandemic have unquestionably saved lives. However, the COVID-19 pandemic has also triggered an unprecedented “infodemic”; the velocity and volume of data production have overwhelmed many key stakeholders such as clinicians and policy makers, as they have been unable to process structured and unstructured data for evidence-based decision making. Solutions that aim to alleviate this data synthesis–related challenge are unable to capture heterogeneous web data in real time for the production of concomitant answers and are not based on the high-quality information in responses to a free-text query.

Objective: The main objective of this project is to build a generic, real-time, continuously updating curation platform that can support the data synthesis and analysis of a scientific literature framework. Our secondary objective is to validate this platform and the curation methodology for COVID-19–related medical literature by expanding the COVID-19 Open Research Dataset via the addition of new, unstructured data.

Methods: To create an infrastructure that addresses our objectives, the PanSurg Collaborative at Imperial College London has developed a unique data pipeline based on a web crawler extraction methodology. This data pipeline uses a novel curation methodology that adopts a human-in-the-loop approach for the characterization of quality, relevance, and key evidence across a range of scientific literature sources.

Results: REDASA (Realtime Data Synthesis and Analysis) is now one of the world’s largest and most up-to-date sources of COVID-19–related evidence; it consists of 104,000 documents. By capturing curators’ critical appraisal methodologies through the discrete labeling and rating of information, REDASA rapidly developed a foundational, pooled, data science data set of over 1400 articles in under 2 weeks. These articles provide COVID-19–related information and represent around 10% of all papers about COVID-19.

Conclusions: This data set can act as ground truth for the future implementation of a live, automated systematic review. The three benefits of REDASA’s design are as follows: (1) it adopts a user-friendly, human-in-the-loop methodology by embedding

XSL•FO
RenderX

an efficient, user-friendly curation platform into a natural language processing search engine; (2) it provides a curated data set in the JavaScript Object Notation format for experienced academic reviewers' critical appraisal choices and decision-making methodologies; and (3) due to the wide scope and depth of its web crawling method, REDASA has already captured one of the world's largest COVID-19–related data corpora for searches and curation.

(J Med Internet Res 2021;23(5):e25714) doi:10.2196/25714

KEYWORDS

structured data synthesis; data science; critical analysis; web crawl data; pipeline; database; literature; research; COVID-19; infodemic; decision making; data; data synthesis; misinformation; infrastructure; methodology

Introduction

Between December 31, 2019, and August 3, 2020, 37,362 papers related to COVID-19 were published on PubMed [1], with Dimensions reporting over 100,743 publications, 1503 policy documents, and 1097 data sets [2]. The speed and scale of the production of published data on COVID-19 across both peer- and nonpeer-reviewed literature presents considerable challenges for stakeholders (eg, policy makers, clinicians, and patients) who must make subjective quality judgements on new data and rapidly synthesize information in order to make optimal, evidence-based decisions. Traditional approaches to data synthesis are unable to keep pace with the rapidly changing information landscape. For example, in the United Kingdom, the National Institute for Health and Care Excellence was unable to publish their initial therapeutic guidance on managing COVID-19 until March 20, 2019 [3]. Ultimately, they modified their methodology for publishing rapid guidance materials on COVID-19 [4]. Moreover, there have been concerns regarding data credibility and the political misuse of information, resulting in the World Health Organization announcing its campaign for discouraging the spread of misinformation [5]. The COVID-19 pandemic highlights the urgent need to prospectively capture, structure, and interpret expansive and complex data sets in real time to support the rapid development of clinical guidance and, most critically, ensure that various key stakeholders can make the best possible evidence-based decisions during an "infodemic."

Previous strategies have attempted to address these challenges by using the concepts of live systematic reviews, which involve the continuous, structured analysis of data that target specific clinical questions [6,7] as well as the clear presentation of such data [8]. However, despite the progress in this field, major obstacles remain in establishing automated data mining frequency, depth, and robustness. Moreover, major barriers exist in the development and validation of machine learning methodologies for the autonomous analysis of heterogeneous clinical data sets.

This paper outlines the methodology of REDASA (Realtime Data Synthesis and Analysis)—a novel, prospective clinical information platform that was developed during the COVID-19 pandemic. It was designed for use across a wide range of data-rich subject areas while keeping application and impact in mind. Our objective was to continuously capture and synthesize both academic and relevant grey literature (eg, news websites, policy documentation, and social media posts) on COVID-19 and to develop a validated data curation approach that could supplement machine learning methodologies and be used as the basis for conducting live systematic reviews.

Methods

Components of REDASA

REDASA was built and deployed on the Amazon Web Service (AWS) cloud. Cloud computing is the on-demand delivery of compute power, database storage, applications, and other information technology resources through a cloud service platform on the internet. A cloud services platform, such as AWS, owns and maintains the network-connected hardware required for these application services. By using the AWS cloud, REDASA components were rapidly designed. REDASA components were integrated and deployed by using AWS tools and the solutions developed by AWS Partners, MirrorWeb, and Cloudwick. These components were comprised of a real-time data extraction pipeline that was implemented by using MirrorWeb's digital archiving technology, a data lake storage repository and workflow orchestration platform (Amorphic) that was developed by Cloudwick, a natural language search engine that was implemented by using Amazon Kendra, and a document curation pathway that was implemented by using Amazon SageMaker Ground Truth.

Real-Time Data Extraction Pipeline

MirrorWeb was used to conduct an exploratory review of the target websites via manual and automatic content detection for informing crawl scoping decisions. Exploratory reviews involve the domain composition analysis of initial web estate archives, which can be produced via multiple methods, including basic link harvesting, Domain Name System lookups, the gathering of URL lists at crawl time (to identify content delivery networks and perform manual verifications), and the inspection of websites. This ensures that (1) the relevant areas of websites are identified and followed by the archive tools and (2) content that can be confidently omitted is avoided. With the adoption of machine learning algorithms, this process can be further assisted by technology.

Scoping decisions can encompass a range of factors. For instance, scoping decisions can ensure that the website crawl stays within the target domain. Further, they can refine the website crawl to only include relevant URL patterns within the domain. For example, if there is a domain.com/coronavirus/ subfolder that contains relevant content, the web crawler will use a series of URL pattern matches and regular expressions to allow or disallow URL strings via exact matching or wildcard pattern matching, thereby containing the crawl to specific areas

XSL•FO
RenderX

of the website. Additionally, scoping decisions can expand the crawl scope to include any outlying content that would be excluded by the refinement rules. Some websites have nonstandard locations for storing assets such as images, stylesheets, and scripts, which are needed to create a high-fidelity representation of the source material. Some websites also contain relevant documents that are located in prescribed location structures outside of the primary target folder.

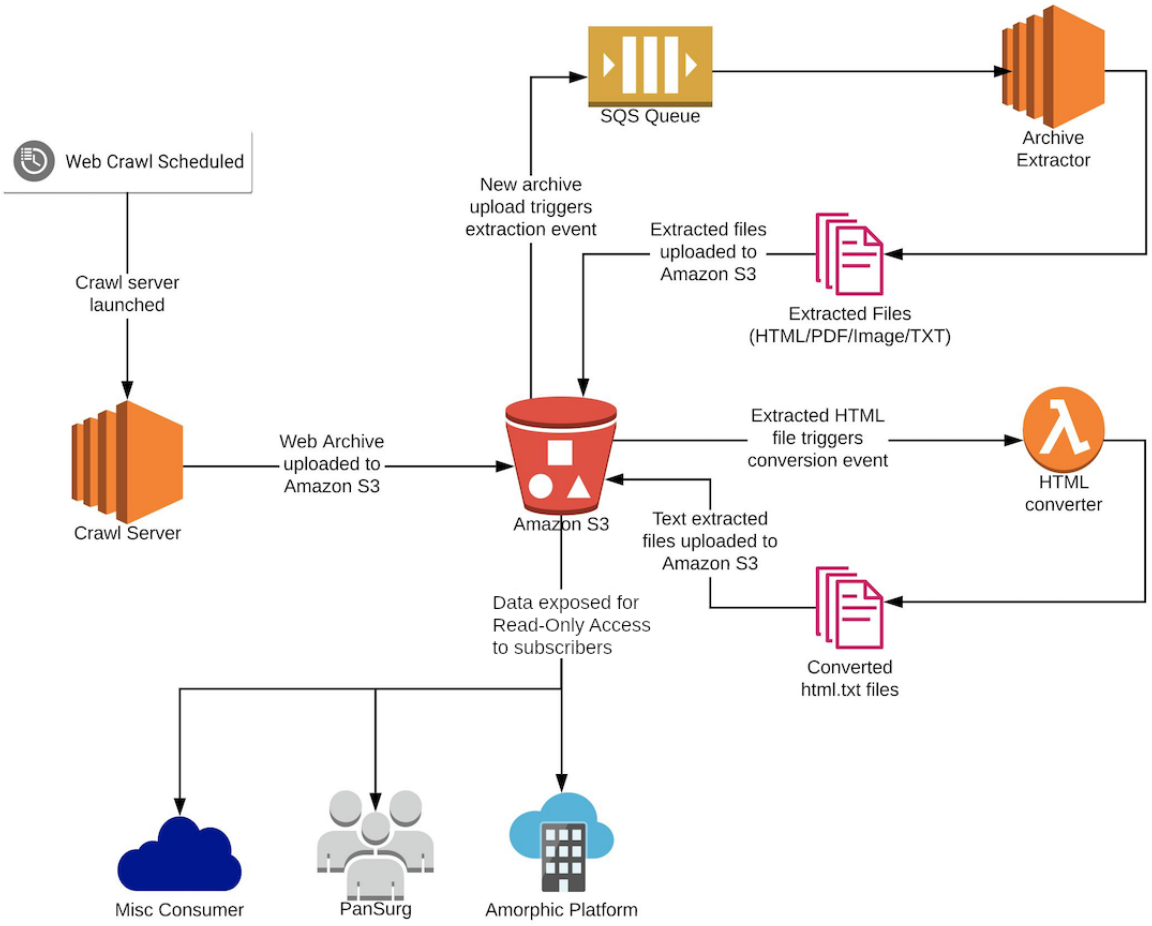
Successful, high-fidelity targeted web crawling has been well documented. However, considerable challenges remain in the development of a qualitative and quantitative method for real-time relevant URL detection [9]. This is because the web ecosystem is a constantly evolving landscape with continuous advancements in available technology; construction techniques; and the consideration of desktop, mobile, and accessible display devices. Furthermore, the sheer number of content management systems that adopt their own proprietary content structures, the advent of single page applications, the prevalence of JavaScript, and people's dependence on asynchronous loading and POST method requests (which returns the same URL for multiple requests) render traditional URL similarity detection a particular challenge. Programmatic links with no human-readable semantic structures and features, such as "Latest News" sections within a web page, can often skew the results of link-based page relevance analyses.

These challenges are exacerbated in the REDASA system, which is required to target data in both academic and nonacademic sources without a guaranteed schema, dialect, or topic. Previously developed methodologies for addressing this issue have been used to apply anchor text similarity scores, content similarity scores, and URL prediction scores (which are based on a set of starting keywords) to seed data [10,11]. These scoring models promote and remove keywords based on the detection of commonalities in discovered crawl data. However, this approach presents several challenges because it relies on good starting URLs that present a reliable and consistent pattern of data. To counter this, REDASA performs downstream content filtering after the crawl is complete in order to eliminate extraneous data, which eliminates the risk of losing vital information that comes with the analysis of a potentially biased set of keywords. Due to REDASA's ability to perform a retrospective analysis of retained data, it will serve as a future platform that can be further enhanced by discovery automation.

In practice, REDASA performs an initial crawl that is launched by using crawl scope definitions that are governed by the aforementioned decision rules. The process for completing a crawl is outlined in Figure 1, which provides an interactive replica of the website content that is accessible to curators. MirrorWeb's Crawl Quality team reviewed the quality of the archive by using a replay engine to create a navigable replica of the target website archive. In addition to clicking the links and manually reviewing the content, an automated link checker was used to recursively spider the archive, identify page and document links in the HTML content, and attempt to open the target URLs in the archive instance in an effort to detect any missing referenced content. If any changes to the scoping rules are needed to make the web crawl more permissive or more restrictive, the scoping rules will be amended as required by using the same, aforementioned scoping principles. This iterative cycle repeats until the crawl is of sufficient quality to be used for human curation and accurate natural language processing (NLP) searches.

XSL•FO
RenderX

Figure 1. The REDASA back-end web crawling and data processing pipeline. REDASA: Realtime Data Synthesis and Analysis; SQS: Simple Queue Service; TXT: text.



Data Lake

Cloudwick's Amorphic platform provides a core REDASA data lake repository and data workflow orchestration service. MirrorWeb data initially lands in the storage layer of Amorphic, which consists of a landing zone. After validation checks are performed, data are moved and stored in a data lake zone and made available for document curation and search index workflows, as described in the following section.

Search Index for Question-Specific Curation Documents

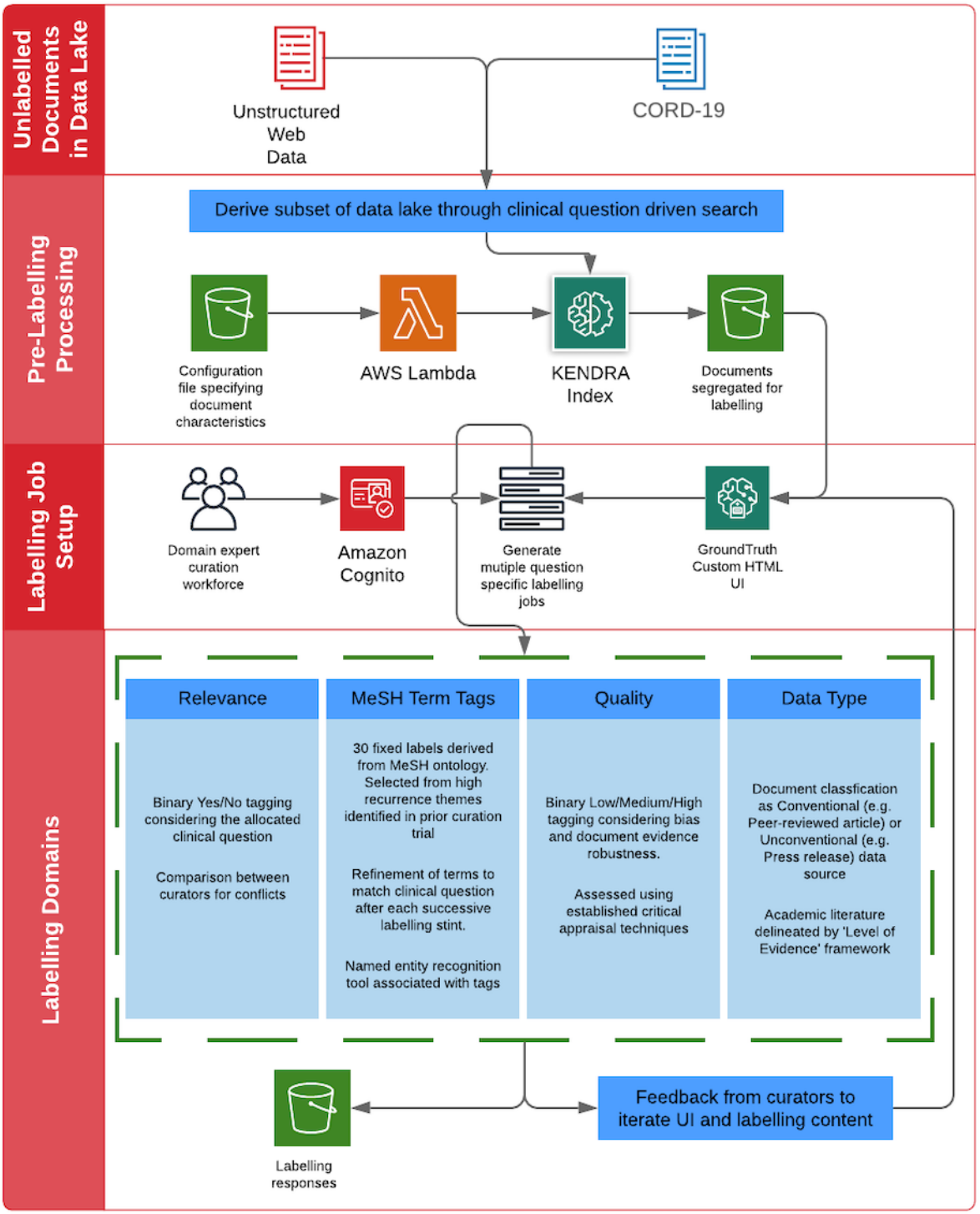
REDASA uses an AWS enterprise search service—Amazon Kendra—to provide search functionality across the entire data lake. Amazon Kendra is an NLP machine learning service that uses deep learning models to understand natural language queries and document content and structures. Amazon Kendra provides support for the following three broad types of questions: (1) factoid questions (who, what, when, and where), which are questions that require fact-based answers that may be returned in the form of a single word or phrase (the precise answer however must be explicitly stated in the ingested text content); (2) descriptive questions, which involves answers that could be a sentence, passage, or an entire document; and (3) keyword searches, wherein the intent and scope of the question may not be clear. In REDASA's question-specific curation model, Amazon Kendra exclusively received factoid questions and used a series of deep learning models to return relevant documents (Figure 2).

The key component of Amazon Kendra is an index. Conceptually, an index is an abstraction that encompasses a set of documents and the underlying hardware and software infrastructure that makes it possible to query documents that use natural language. Aside from its actual content, each document may include some associated metadata (eg, the source of the document, the document's logical unit group, etc). Users can specify custom metadata fields to suit their needs. These metadata tags are accessible through the Amazon Kendra query application programming interface.

A Kendra index may consist of one or more data sources, and a data source is a logical unit of documents. For REDASA, data source file types were limited to plain text, HTML, and PDF. Compared to other file types, these better integrate with our curation platform and allow for consistent labeling outputs when performing named entity recognition (NER).

XSL•FO
RenderX

Figure 2. Integrated workflow of the search index and data curation pipeline for a variety of high-impact areas with and without consensus among the scientific community in different countries and health authority bodies. AWS: Amazon Web Service; CORD-19: COVID-19 Open Research Dataset; MeSH: Medical Subject Headings; UI: user interface.



Document Curation

Document curation was implemented by using the custom workflows in Amazon SageMaker Ground Truth, which is a data labeling service that is used to build training data sets for machine learning workflows. REDASA uses a question-based curation approach. PanSurg investigators posed a series of COVID-19–related key questions to the search index (Textbox 1). These questions were chosen to obtain answers, and we were able to validate the quality of the data lake and the adequacy of REDASA's data mining depth with our curation relevance metric.

XSL•FO
RenderX

Textbox 1. COVID-19–related natural language processing queries that were posed to the REDASA (Realtime Data Synthesis and Analysis) search index to develop a question-specific curation methodology.

Queries in natural language

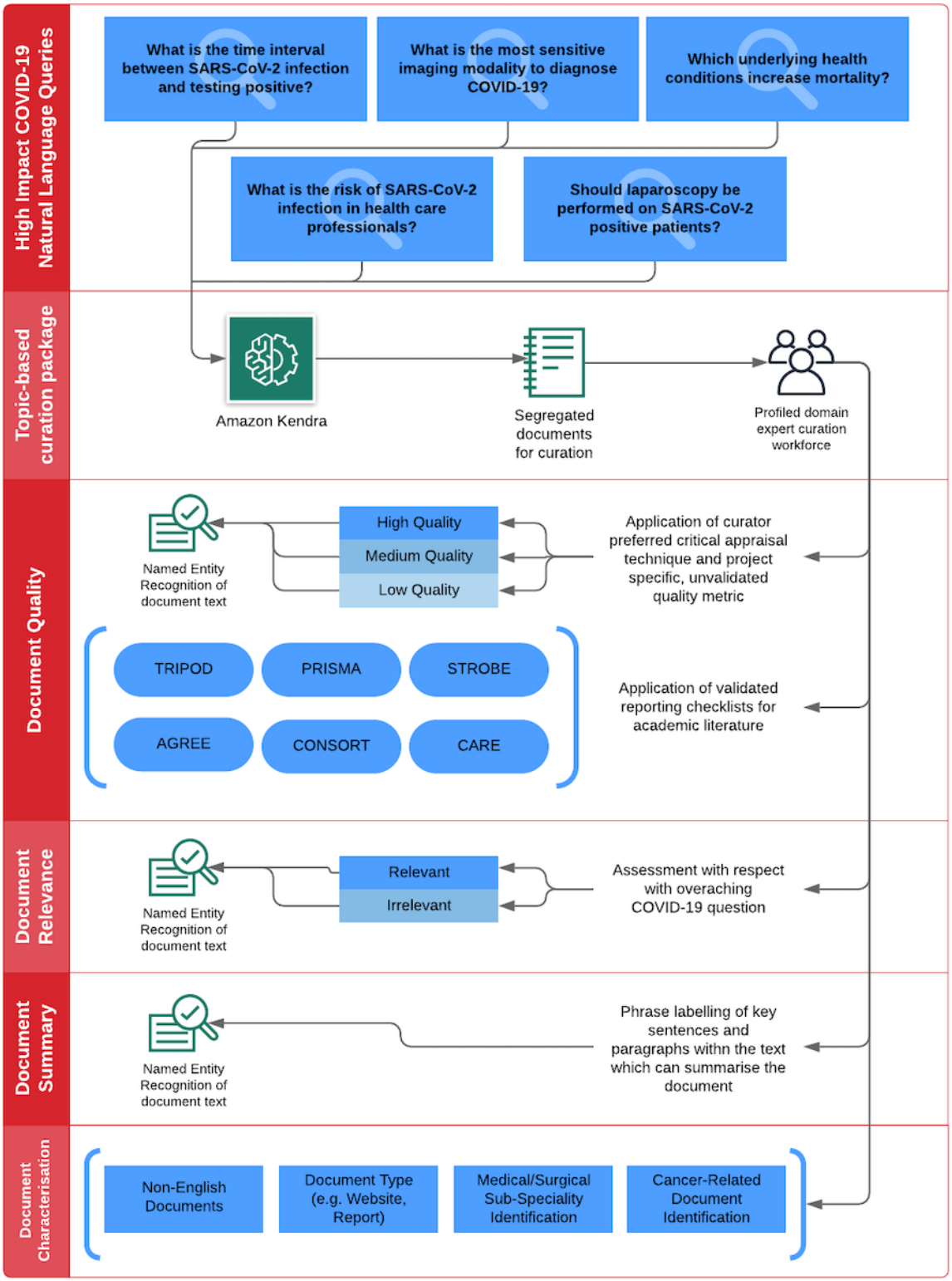
1. What is the time interval between SARS-CoV-2 infection and testing positive?
2. What is the most sensitive imaging modality to diagnose COVID-19?
3. Which underlying health conditions increase mortality?
4. What is the risk of SARS-CoV-2 infection in health care professionals?
5. Should laparoscopy be performed on SARS-CoV-2 positive patients?
6. What is the estimated global economic impact of COVID-19?
7. How effective are masks at minimizing the transmission of COVID-19 in public?
8. What is the evidence for COVID-19 mutations and how many subtypes are responsible for the pandemic?
9. Does a positive SARS-CoV-2 antibody test mean an individual is immune to further COVID-19 infections?
10. Is COVID-19 airborne transmitted?
11. Can asymptomatically infected individuals transmit COVID-19?
12. What is the evidence for 1-meter and 2-meter separations for social distancing?
13. What has the evidence-base been for lockdown compared to no lockdown during this COVID-19 pandemic?
14. Is universal preoperative testing for SARS-CoV-2 beneficial compared to selective testing?
15. Can individuals be reinfected with SARS-CoV-2?

The REDASA search index provides a list of selected documents, which are randomly provisioned to curators for labeling via Amazon SageMaker Ground Truth. This allows them to assess documents' relevance and quality in relation to the original query and further categorize the data based on the labels described in Figure 3. Reflecting the living nature of the REDASA platform, queries were adapted in accordance with the knowledge priorities of different phases of the pandemic. For example, in stint 1 of curation (February 6 to September 6, 2020), which was scheduled during the peak of the UK COVID-19 outbreak and when uncertainties regarding best practices for screening and management planning were rife, questions 1-5 were posed to REDASA. In contrast, stint 2 of curation was performed during the nationwide lockdown relaxation period and the public health transition for minimizing the risk of a second wave of COVID-19. Consequently, questions 6-15 focused upon themes such as reinfection, transmission mitigation, and the global impact of the pandemic.

The relevance of articles in relation to the query that was posed to the search index was a subjective binary measure (ie, irrelevant or relevant). This assessment was paired with NER labels, which enabled curators to highlight phrases and paragraphs that indicated the relevance of articles.

XSL•FO
RenderX

Figure 3. Curation labels for generating document metadata. AGREE: Appraisal of Guidelines for Research and Evaluation; CARE: Case Reports; CONSORT: Consolidated Standards of Reporting Trials; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

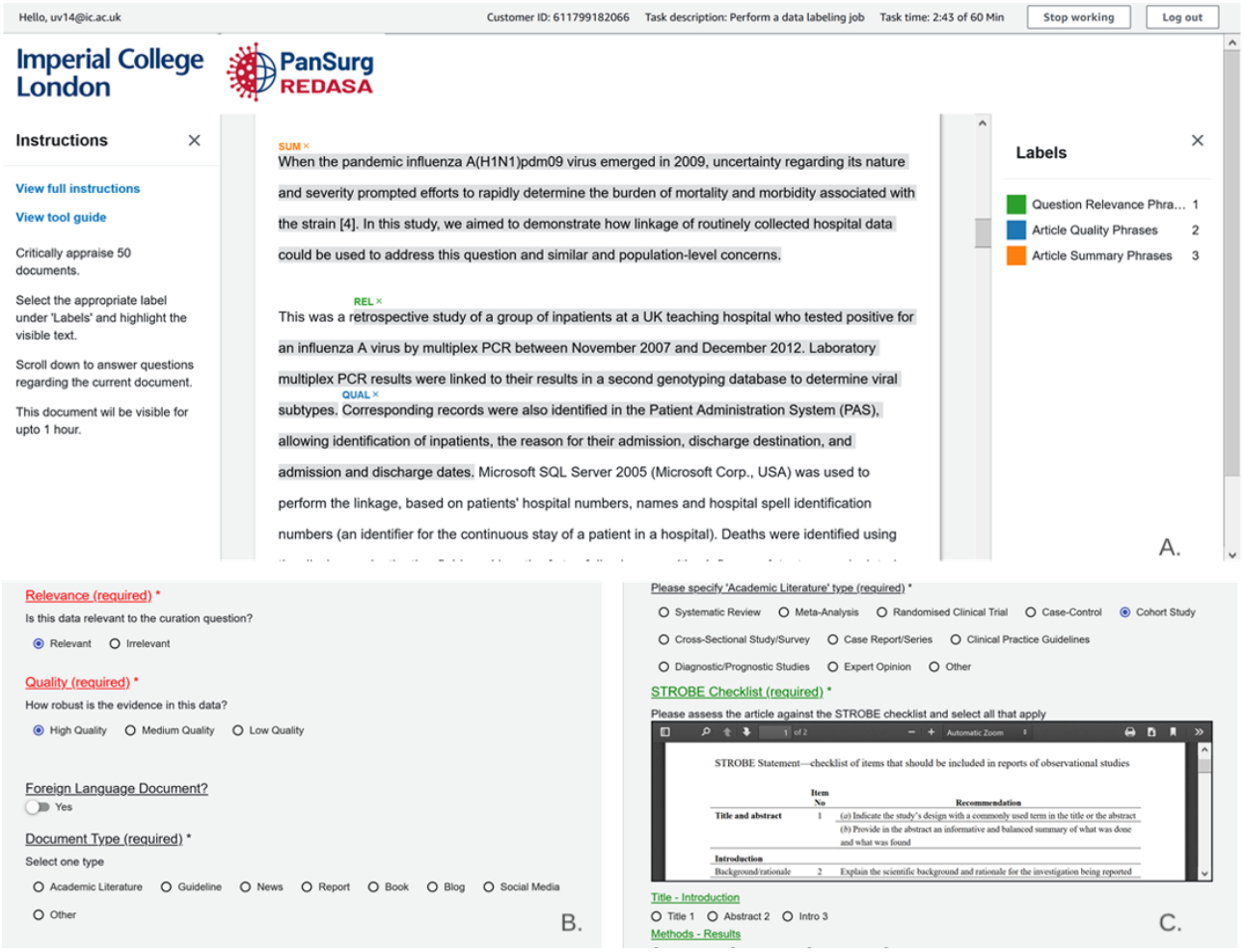


The quality of the academic literature was assessed via a 3-stage process. First, we ascertained the study type, and this allowed us to assign an evidence rating level (the levels proposed by the Oxford Centre for Evidence-Based Medicine [12]). Second, we invited curators to provide an independent, subjective rating of an article's quality by using their own critical appraisal methodology and assign 1 of the 3 following binary ratings: low, medium, or high. Third, akin to the relevance metric, NER annotation was made available to curators and correlated with their low, medium, or high ratings. Depending on the type of academic literature that curators were assessing, curators were automatically given (through their user interface) the relevant EQUATOR (Enhancing the Quality and Transparency of Health Research) checklist for quantitative quality assessment [13].

XSL•FO
RenderX

For example, if the document that curators were assessing was a systematic review, they were automatically able to assess the article against the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Figure 4).

Figure 4. (A) A document with curation user interface labels (the NER of quality, relevance, and summary phrases). (B) Binary labels for classifying documents and correlating them to NER responses. (C) Embedded reporting checklists for document assessment, which were provided based on the selected academic study type. NER: named entity recognition; REDASA: Realtime Data Synthesis and Analysis; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology.



Collectively, the relevance and quality metrics' utility was threefold. First, they enabled us to capture data on curators' decision-making and critical appraisal processes. Second, they minimized the number of undesirable irrelevant documents, which allowed us to implement a human-in-the-loop optimization methodology for the search index. Third, they allowed us to perform multiple curator passes on a single document, assess for labeling response conflicts, and ascertain the article factors responsible for any disparities. To obtain further data, we allowed our curators to assess the risk of bias in the articles by using the bias metrics designed by Cochrane [14]. The results from this novel curation process were intended to (1) act as ground truth for data science models that aimed to facilitate the future semiautomation or full automation of article screening, and (2) be used for a structured assessment of evidence quality.

This question-specific approach was selected over the more traditional approach of randomly sectioning data to help us preserve the relevance metric for specific questions. This factor would have otherwise been more challenging to implement and capture. Further, this metric is key to future work streams that determine the relevance of specific articles for inclusion in an automated systematic review.

Curation Methodology

Structured and Unstructured Data Lake

Our proof-of-concept analysis for data mining during predefined time periods was feasible. In this iteration of REDASA, a 1-week time period was chosen to enable the capturing of the highest possible number of new data points at the lowest mining frequency, thus minimizing the computing costs of both COVID-19–structured and unstructured data sources. In this paper, we only present a text-based analysis of the data set. In the future, we intend to assess structured, quantitative data from target data sources.

Unstructured Web Crawl Data

Textual information was extracted from a precurated set of internet sites. A range of frequently accessed but disparate data types that are typically used by frontline clinicians and policy makers were extracted. These included high-quality journal websites and portals containing COVID-19–related literature;

XSL•FO
RenderX

medical and surgical society guidance web pages; and guideline repositories from local, governmental, and international public health bodies. By being able to dynamically capture data and automatically obtain updates from sources, this type of data mining demonstrates the power of REDASA in terms of amalgamating qualitative and quantitative insights for generating future reports. Each website was independently assessed and evaluated for inclusion into the REDASA data lake by clinical (n=4) and data science (n=2) reviewers. Disagreements were resolved through consensus. These sources were selected in accordance with criteria for including usable content and determining the reliability and breadth of target topics and categories pertaining to COVID-19. To systematize the data lake prior to data ingestion, sources were categorized into the following broad groups, which were independently defined a priori by the three members of the research team based on the source of the original data: all (miscellaneous), critical care, medical care, surgical care, drug development and pharmacological therapy, mental health, risk, translational research, biological sciences, engineering, and policy. The content of the data from each of the sources was screened by these three independent members of the research team. If disagreements regarding categorization occurred, a meeting was conducted. Unanimous agreement was sought prior to final categorization.

Structured Data From the COVID-19 Open Research Dataset

The White House and a coalition of leading research groups developed the COVID-19 Open Research Dataset (CORD-19) [15]. The CORD-19 is a data set that contains over 157,000 scholarly articles, including over 75,000 full-text articles regarding COVID-19, SARS-CoV-2, and related coronaviruses. This freely available data set was provided to the global research community of Kaggle and was used as a test data source when initially developing the REDASA infrastructure.

Collectively, these assimilated data sources make REDASA one of the world's largest and most contemporaneous COVID-19–related evidence sources, consisting of 104,000 documents.

Data Availability

The curation labels can be found on GitHub [16].

Results

Curation Results

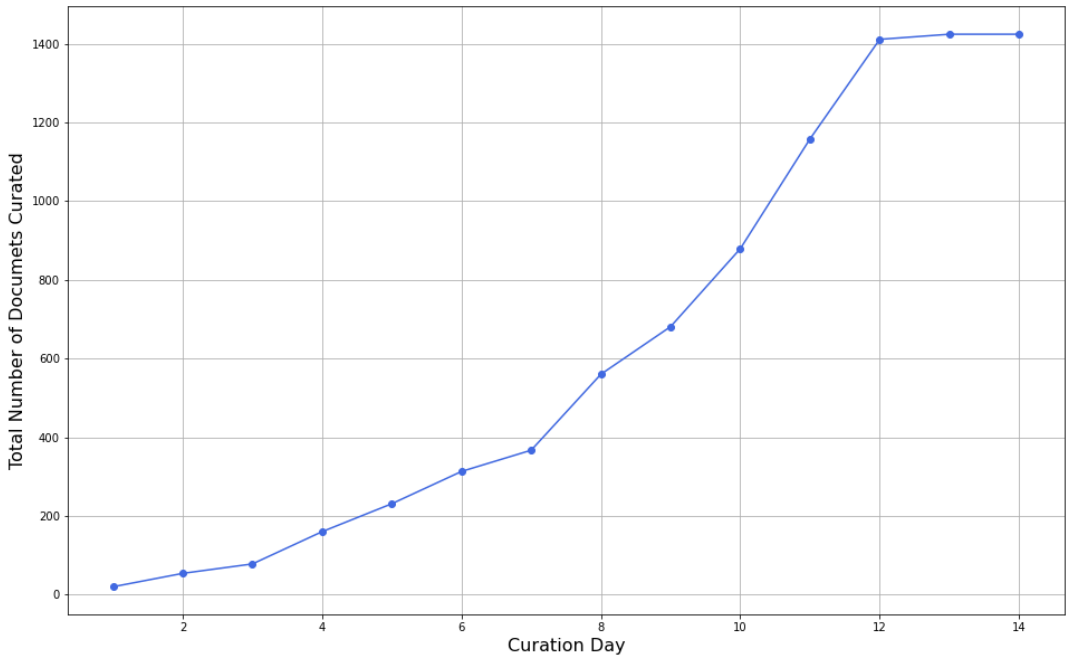
The first-pass document curation responses from 42 curators underwent the preliminary analysis of the domains of relevance to free-text queries and information quality. These data were collected over 2 time-limited curation stints. Both stints were 1 week in duration. We obtained a total of 1424 documents pertaining to 15 different COVID-19–related queries (around 99 text documents derived per query from the search index for the structured and unstructured data lake), and an average of 42 documents per query were assessed by 1 curator. Our aim was for each document to be assessed by at least 2 different reviewers so that we could assess intercurator variability and identify the reasons for discrepancies in evidence quality verdicts.

Each curator was profiled to ascertain their academic or clinical backgrounds. This was initially performed for the vetting and quality control of curation responses. These data will also be used to further discriminate between labeling responses that are based on critical appraisal expertise and to assign weights to curation responses. To date, the REDASA project's international curator community includes people from 9 different countries, and the project is supported by medical and surgical health care professionals who range from senior consultant vascular surgeons in Italy to general practice physicians who are involved in community and public health decision making in the Philippines. Data curators were recruited by invitation through the PanSurg email subscriber list and by open invitation via the @pansurg Twitter account. Curators were included if they had a medical degree or a higher degree in science-related fields. Data curators were asked to state their interest and were verified by direct contact. The number of data curation responses for REDASA, which exponentially rose between our two stints as more curators were onboarded (stint 1: n=12; stint 2: n=42; Figure 5), was indicative of an efficient, novel methodology for the digital, community-based peer review of literature by domain experts. This was exemplified by the ability of some of our experienced curators, who were able review over 100 scientific documents of varying length in as little as 3 days. Furthermore, with curators' 100% follow-through rate between stint 1 and stint 2, our curation model suggests that, when combined with our simplified critical appraisal interface, the peer review of literature at scale is viable and sustainable.

In total, 70.9% (1009/1424) of the pool of curated articles was composed of peer-reviewed, traditional, academic literature; the remainder consisted of web crawl–derived data, including governmental policies and reports from professional bodies. Based on the subset of the 900 academic literature documents that were curated, the most common study type encountered was systematic reviews (98/1009, 9.7%). The least common study type in the data lake was randomized controlled trials (RCTs; 3/1009, 0.3%). Nonsystematic reviews (eg, rapid reviews and comprehensive reviews) were not given the systematic review label to avoid inappropriate assessments against the PRISMA checklist. Such outlier academic literature types were aggregated into the miscellaneous category, which included 427 documents that were assessed solely against the curator-reported binary quality ratings.

XSL•FO
RenderX

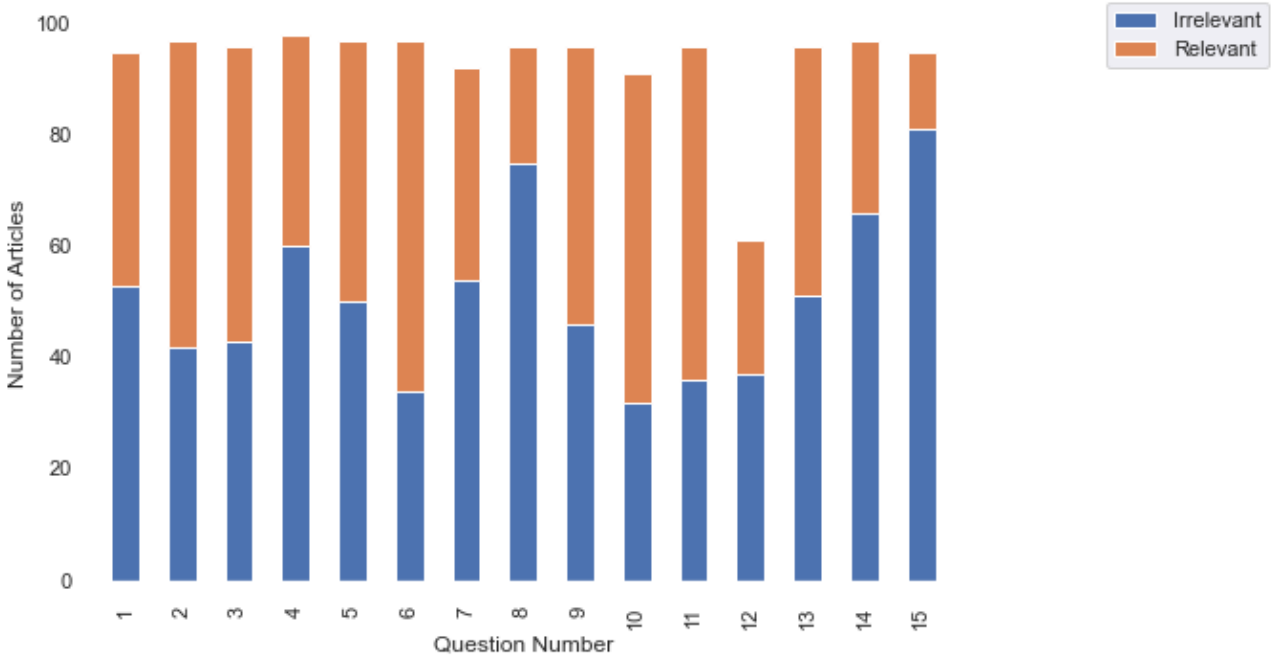
Figure 5. Rate of COVID-19–related scientific literature curation over 2 weeks. This was associated with the growth of the number of curators, which plateaued on day 13. This was when all of the documents available for curation were assessed before the end of stint 2.



Articles' Relevance to Queries

The relevance metric that curators used provided insight into the performance of the search index in terms of providing cogent and useful document results associated with the 15 COVID-19–related queries (Textbox 1). Overall, 50.49% (719/1424) of articles were considered relevant to their respective queries. When observing the question bank, this variance in article relevance (which was based on the search index) was reflected by the lack of consistency in the ratio of the number of relevant articles to the number of irrelevant articles (Figure 6). These data can be used to provide feedback for the search index with regard to the optimization of provided results.

Figure 6. Curators' responses determined the relevance of documents to search index queries. Responses were matched to the query number.



XSL•FO
RenderX

Critical Appraisal of Article Quality

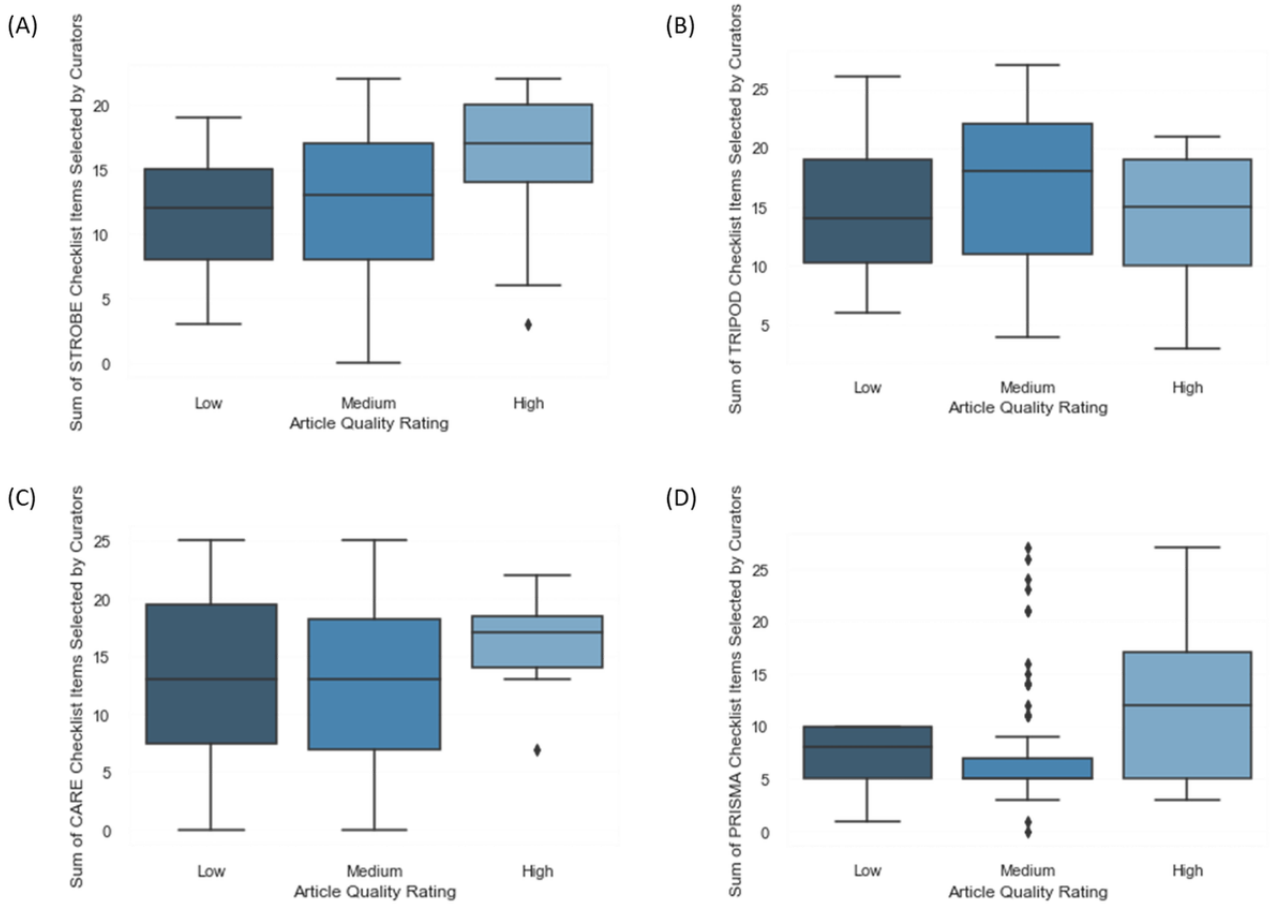
The uneven distribution of academic study types that have been curated thus far precludes the interpretation of results for quantitative reporting checklist responses based on the qualitative rating system (low, medium, and high) for RCTs (CONSORT [Consolidated Standards of Reporting Trials] checklist) and clinical guidelines (AGREE [Appraisal of Guidelines for Research and Evaluation] checklist). These studies were poorly represented in this run of analysis.

Quality was quantified by ascertaining the sum of the number of EQUATOR Network–derived checklist items that were fulfilled by each document. Hence, documents with methods and results that aligned more closely to their respective reporting checklist were scored higher and deemed to be of greater quality. This outcome was compared to the curators' subjective ratings for diagnostic and prognostic studies (TRIPOD [Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis] checklist), case reports and series (CARE [Case Reports] checklist), case-control studies (STROBE [Strengthening the Reporting of Observational Studies in Epidemiology] checklist) and meta-analyses and systematic reviews (PRISMA checklist) (Figure 7). Notably, the subjective quality rating was assigned prior to assessment by using the checklist under our curation protocol to mitigate observer bias.

Based on our independently assessed and subjective quality metric, our preliminary results suggested that more than 50% (726/1424, 50.98%) of the documents derived from the REDASA data pipeline were of medium quality, and 13.55% (193/1424) were deemed high quality. Thus, during data aggregation, 64.53% (919/1424) of the documents derived from our data lake for curation were of a sufficient quality to inform their professional decision-making processes or could be used as reliable sources of information.

With regard to the TRIPOD-relevant (score: mean 15.6, SD 6.9) and STROBE-relevant (score: mean 13.2, SD 5.8) study types, there was some correlation between curators' subjective assessments of the articles that had low-to-high ratings and underwent the validated checklist–based quality assessment. However, with regard to the CARE-relevant (score: mean 13.8, SD 6.8) and PRISMA-relevant (score: mean 8.4, SD 6.1) study types, there was substantial variance in the number of checklist items that were selected for each quality rating, thus indicating an apparent dissociation between these two metrics. Further data collection and the comparison of intercurator responses for the removal of outliers will provide clarity on the role of subjective quality ratings versus the role of validated reporting checklists as a surrogate marker of evidence quality.

Figure 7. Relationship between the low, medium and, high curator-determined quality ratings of (A) case-control studies, (B) diagnostic and prognostic studies, (C) case reports and series, and (D) meta-analyses and systematic reviews and their respective reporting checklist scores. CARE: Case Reports; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.



XSL•FO
RenderX

Discussion

Globally, there are several efforts underway for systematically accruing COVID-19–related data and specifically querying these data and output-relevant literature [17]. These efforts, such as Google's COVID-19 Research Explorer [18], COVIDask [19], and COVIDScholar [20], are fundamentally based on NLP searches. Additionally, Google's solution incorporates the use of follow-up queries associated with a primary question to obtain more focused results. These efforts universally incorporate the CORD-19 and intuitively present output data. Nevertheless, these approaches do not account for the quality of the data source, which is left to the interpretation of the user. Other efforts, such as SciSight [21] and COVID-19 Primer [22], structure data from the CORD-19 into themes (author and engagement), thereby allowing users to make links and review specific topics, albeit without a natural language interface for answering specific questions.

The crucial difference between REDASA and the aforementioned platforms is threefold. First, REDASA adopts a human-in-the-loop methodology by embedding an efficient, user-friendly curation platform into an NLP search engine. REDASA can iteratively refine its search outputs at scale, particularly in the domains of the relevance and quality of data sources. This can ultimately contribute to a fact-checking function for conducting a reliable assessment of the utility of an article [23]. Second, it provides a curated data set in the JavaScript Object Notation format for experienced academic reviewers' critical appraisal choices and decision-making methodologies. These data on the peer-review process provide a unique framework for modelling, quantifying, and ultimately automating the evidence quality assurance process and are unavailable elsewhere. Finally, due to the wide scope and depth of REDASA's web crawling methodology, REDASA has already captured one of the world's largest COVID-19–related data corpora for searches and curation. Our aim is to make these crucial data freely available and ensure that they are continuously updated to allow for rapid review and dissemination during and beyond the evolving pandemic.

For the long-term goal of conducting a semisupervised, live systematic review of data, several limitations and challenges need to be overcome. Our curation methodology resulted in a high turnover rate for the assessment of data. However, there was still variability in curator output, which was secondary to the variability in curators' subjective critical appraisals. In this project, we relied on the prescreening of curators, which was conducted via academic portfolio screening and assessments for relevant literature review experience. This crucial quality control approach needs to be further developed to fully validate and enhance the accuracy of our curation methodology. A limitation of our preliminary data analysis was the qualitative, summative comparison of the EQUATOR checklist ratings to our quality ratings. This was due to the subcomponents of the used EQUATOR checklists, which did not use equal metrics for article quality, and the nonexhaustive quality criteria captured by these tools. Hence, future studies are needed to validate our quality ratings and identify a reliable metric for quality that is applicable across the academic and nonacademic literature captured by REDASA. In addition to ensuring the consistency of quality ratings, sustained curation work is required to ensure that the corpus includes greater numbers of studies across all designs and methodologies—specifically, RCTs (if available)—to ensure that the corpus is truly representative of data under examination.

Our framework has demonstrated proof-of-concept that by combining the discovery and ingestion pipeline, data lake repository, human curation platform, and NLP semantic search index of REDASA, it can provide curated responses to questions that are posed in natural language in the short term. In the long term however, based on the data insights that progressively validated the critical appraisal of our curation methodology, the ambition of REDASA is to conduct live systematic reviews by using semisupervised machine learning techniques to rapidly return high-quality, relevant evidence in response to queries for any discipline experiencing an "infodemic," such as cancer or cardiovascular disease.

Acknowledgments

We would like to thank the team at AWS Healthcare for their support and help throughout the development of this project and Simon Russell and Steve Neufville for their technical and project management support, respectively. We would also like to thank the PanSurg Collaborative for their support during the project. We thank David Hilman for his technical support with the final stages of this paper. We also thank Elizabeth Burgnon for carefully proofreading the final paper. This work was supported by Defence and Security Accelerator (grant ACC2015551), the Digital Surgery Intelligent Operating Room Grant, the National Institute for Health Research Long-limb Gastric Bypass RCT Study, the Jon Moulton Charitable Trust Diabetes Bariatric Surgery Grant, the National Institute for Health Research (grant II-OL-1116-10027), the National Institutes of Health (grant R01-CA204403-01A1), Horizon H2020 (ITN GROWTH), and the Imperial Biomedical Research Centre.

The members of the PanSurg REDASA Curators are Ademola Adeyeye (Afe Babalola University), Ahmed Ezzat (Imperial NHS Trust), Alberto Porcu (Università degli Studi di Sassari, Italia), Alexander Walmsley (University of Liverpool), Ali Farsi (King Abdulaziz University, Jeddah, Saudi Arabia), Alison Faye O Chan (Institute of Biology, College of Science, University of the Philippines Diliman), Aminah Abdul Razzack (Dr NTR University of Health Sciences), Andee Dzulkarnaen Zakaria (School of Medical Sciences, Universiti Sains Malaysia), Andrew Yiu (Guy's and St Thomas' NHS Foundation Trust), Antonios Soliman (Assiut University Hospitals), Ariana Axiaq (Queen's University Belfast), Avinash Aujayeb (Northumbria HealthCare NHS Foundation Trust), Catherine Dominic (Barts and the London School of Medicine, Queen Mary University of London), Eduarda

XSL•FO
RenderX

Sá-Marta (Faculty of Medicine, University of Coimbra), Eunice F Nolasco (Independent Researcher), Jessamine Edith S Ferrer (Blessed Family Doctors General Hospital), Jonathan Anthony Kat (University of Manchester), Josephine Holt (Chelsea and Westminster NHS Trust), Kamal Awad (Faculty of Medicine, Zagazig University, Egypt), Kirk Chalmers (University of Liverpool), Mina Ragheb (University of Liverpool School of Medicine), Muhammad Khawar Sana (King Edward Medical University), Niraj Sandeep Kumar (UCL Medical School), Roland Amoah (Newcastle upon Tyne Hospitals Trust), Semra Demirli Atici (University of Health Sciences Tepecik Training and Research Hospital), Shane Charles (San Fernando General Hospital), Sunnia Ahmed (Independent Researcher), Teresa Perra (Università degli Studi di Sassari, Italia), Tricia Hui Chyi TAY (Manchester Medical School, University of Manchester), Ubaid Ullah (Medical Emergency Resilience Foundation Pakistan), Zara Ahmed (King's College London), and Zun Zheng Ong (University of Nottingham).

Authors' Contributions

JK, GM, and MH were responsible for project conception and overall management. The development of the data pipeline and data lake structure was steered by KS, EF, HS and DB. The implementation of the MirrorWeb web crawling technology was led by KS, who received guidance from UV, SR, PB, and OS. The initial development of the curation platform was performed by EF, and subsequent optimization and tailoring for REDASA use was conducted by CL, UV, SR, and OS. The coordination and development of the data curation methodology was performed by UV, SR, PB, and OS. Data analysis and manuscript composition was performed by UV and SR, who received input from JK, GM, OS, SP, JC, KS, EF, and MH. All authors reviewed and contributed to the manuscript. We would like to thank the PanSurg REDASA Curators for their support on the project.

Conflicts of Interest

GM received equity from Medical iSight (Augmented Reality). 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 One Welbeck Hospital. JK provides consultations for Verb robotics, Safeheal, YSOPIA bioscience, and Universal Diagnostics (UDX). JK also received equity from Mangetoo (teledietetics), 1 Welbeck Day Surgery (Hospital), 1 World medical (Personal Protective Equipment), and Medical iSight (Augmented Reality). The other authors have no conflicts of interest to declare for this paper.

References

1. LitCovid. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/research/coronavirus/ [accessed 2021-04-12]
2. COVID-19 research update: How many pandemic papers have been published? Nature Index. URL: https://www.natureindex.com/news-blog/how-coronavirus-is-changing-research-practices-and-publishing [accessed 2020-10-01]
3. COVID-19 rapid guideline: critical care in adults. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng159 [accessed 2020-10-01]
4. Interim process and methods for developing rapid guidelines on COVID-19. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/process/pmg35/chapter/scoping [accessed 2020-10-01]
5. Countering Missinformation about COVID-19. World Health Organization. URL: https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19 [accessed 2020-10-01]
6. Maguire BJ, Guérin PJ. A living systematic review protocol for COVID-19 clinical trial registrations. Wellcome Open Res 2020 Apr 02;5:60 [FREE Full text] [doi: 10.12688/wellcomeopenres.15821.1] [Medline: 32292826]
7. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med 2014 Feb;11(2):e1001603 [FREE Full text] [doi: 10.1371/journal.pmed.1001603] [Medline: 24558353]
8. Lenaerts G, Bekkering GE, Goossens M, De Coninck L, Delvaux N, Cordyn S, et al. Tools to Assess the Trustworthiness of Evidence-Based Point-of-Care Information for Health Care Professionals: Systematic Review. J Med Internet Res 2020 Jan 17;22(1):e15415 [FREE Full text] [doi: 10.2196/15415] [Medline: 31951213]
9. Olston C, Najork M. Web Crawling. Boston, United States: Now Publishers Inc; 2010.
10. Rungsawang A, Angkawattanawit N. Learnable topic-specific web crawler. Journal of Network and Computer Applications 2005 Apr;28(2):97-114. [doi: 10.1016/j.jnca.2004.01.001]
11. Zowalla R, Wetter T, Pfeifer D. Crawling the German Health Web: Exploratory Study and Graph Analysis. J Med Internet Res 2020 Jul 24;22(7):e17853 [FREE Full text] [doi: 10.2196/17853] [Medline: 32706701]
12. Oxford Centre for Evidence-Based Medicine: Levels of Evidence. Centre for Evidence-Based Medicine. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 [accessed 2021-04-22]
13. Enhancing the QUAlity and Transparency Of health Research. The Equator Network. URL: https://www.equator-network.org/ [accessed 2021-04-22]
14. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Glasgow, Scotland: The Cochrane Collaboration and John Wiley & Sons; Nov 20, 2019.

XSL•FO
RenderX

15. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The Covid-19 Open Research Dataset. ArXiv Preprint posted online on April 22, 2020. [Medline: 32510522]
16. PanSurg/redasa-sample-data. GitHub. URL: https://github.com/PanSurg/redasa-sample-data [accessed 2021-04-26]
17. Hutson M. Artificial-intelligence tools aim to tame the coronavirus literature. Nature 2020 Jun 9 [FREE Full text] [doi: 10.1038/d41586-020-01733-7]
18. COVID-19 Research Explorer. Google. URL: https://covid19-research-explorer.appspot.com/ [accessed 2021-04-22]
19. COVIDAsk. DMIS Lab, Korea University. URL: https://covidask.korea.ac.kr/ [accessed 2020-10-01]
20. COVIDScholar. COVIDScholar. URL: https://covidscholar.org/ [accessed 2021-04-22]
21. SciSight - About. SciSight. URL: https://scisight.apps.allenai.org/ [accessed 2021-04-22]
22. COVID-19 Primer. Primer.ai. URL: https://covid19primer.com/dashboard [accessed 2021-04-22]
23. Wadden D, Lin S, Lo K, Wang LL, van Zuylen M, Cohan A, et al. Fact or fiction: Verifying scientific claims. arXiv Preprint posted online on April 30, 2020. [doi: 10.18653/v1/2020.emnlp-main.609]

Abbreviations

AGREE: Appraisal of Guidelines for Research and Evaluation
AWS: Amazon Web Service
CARE: Case Reports
CONSORT: Consolidated Standards of Reporting Trials
CORD-19: COVID-19 Open Research Dataset
EQUATOR: Enhancing the Quality and Transparency of Health Research
NER: named entity recognition
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
REDASA: Realtime Data Synthesis and Analysis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by C Basch; submitted 12.11.20; peer-reviewed by S Liu, B Sousa-Pinto; comments to author 11.12.20; revised version received 30.12.20; accepted 03.04.21; published 06.05.21.

Please cite as:
Vaghela U, Rabinowicz S, Bratsos P, Martin G, Fritzilas E, Markar S, Purkayastha S, Stringer K, Singh H, Llewellyn C, Dutta D, Clarke JM, Howard M, PanSurg REDASA Curators, Serban O, Kinross J
Using a Secure, Continually Updating, Web Source Processing Pipeline to Support the Real-Time Data Synthesis and Analysis of Scientific Literature: Development and Validation Study
J Med Internet Res 2021;23(5):e25714
URL: https://www.jmir.org/2021/5/e25714
doi:10.2196/25714
PMID:33835932

©Uddhav Vaghela, Simon Rabinowicz, Paris Bratsos, Guy Martin, Epameinondas Fritzilas, Sheraz Markar, Sanjay Purkayastha, Karl Stringer, Harshdeep Singh, Charlie Llewellyn, Debabrata Dutta, Jonathan M Clarke, Matthew Howard, PanSurg REDASA Curators, Ovidiu Serban, James Kinross. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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

Associations Between Health Literacy, eHealth Literacy, and COVID-19–Related Health Behaviors Among Chinese College Students: Cross-sectional Online Study

Shaojie Li[1], MD; Guanghui Cui[2], BSc; Atipatsa Chiwanda Kaminga[3,4], PhD; Sixiang Cheng[1], PhD; Huilan Xu[1], PhD

[1]Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, Changsha, China

[2]School of Acupuncture and Tuina, Shandong University of Traditional Chinese Medicine, Jinan, China

[3]Department of Mathematics and Statistics, Mzuzu University, Luwinga, Malawi

[4]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

Corresponding Author:
Huilan Xu, PhD
Department of Social Medicine and Health Management
Xiangya School of Public Health
Central South University
238 Shangmayuanling Lane
Changsha, 410078
China
Phone: 86 13637480958
Email: xhl_csu@163.com

Abstract

Background: During the COVID-19 pandemic, the internet has significantly spread information, providing people with knowledge and advice about health protection regarding COVID-19. While a previous study demonstrated that health and eHealth literacy are related to COVID-19 prevention behaviors, few studies have focused on the relationship between health literacy, eHealth literacy, and COVID-19–related health behaviors. The latter includes not only preventative behaviors but also conventional health behaviors.

Objective: The objective of this study was to develop and verify a COVID-19–related health behavior questionnaire, explore its status and structure, and examine the associations between these behaviors and participants' health literacy and eHealth literacy.

Methods: A snowball sampling method was adopted to recruit participants to complete anonymous cross-sectional questionnaire surveys online that assessed sociodemographic information, self-reported coronavirus knowledge, health literacy, eHealth literacy, and COVID-19–related health behaviors.

Results: Of 1873 college students who were recruited, 781 (41.7%) had adequate health literacy; the mean eHealth literacy score was 30.16 (SD 6.31). The COVID-19–related health behavior questionnaire presented a two-factor structure—COVID-19–specific precautionary behaviors and conventional health behaviors—with satisfactory fit indices and internal consistency (Cronbach α=.79). The mean score of COVID-19–related health behaviors was 53.77 (SD 8.03), and scores differed significantly (P<.05) with respect to residence, college year, academic major, family economic level, self-reported health status, having a family member or friend infected with coronavirus, and health literacy level. Linear regression analysis showed that health literacy and eHealth literacy were positively associated with COVID-19–specific precautionary behaviors ($\beta_{\text{health literacy}}$=.149, $\beta_{\text{eHealth literacy}}$=.368; P<.001) and conventional health behaviors ($\beta_{\text{health literacy}}$=.219, $\beta_{\text{eHealth literacy}}$=.277; P<.001).

Conclusions: The COVID-19–related health behavior questionnaire was a valid and reliable measure for assessing health behaviors during the pandemic. College students with higher health literacy and eHealth literacy can more actively adopt COVID-19–related health behaviors. Additionally, compared to health literacy, eHealth literacy is more closely related to COVID-19–related health behaviors. Public intervention measures based on health and eHealth literacy are required to promote COVID-19–related health behaviors during the pandemic, which may be helpful to reduce the risk of COVID-19 infection among college students.

(J Med Internet Res 2021;23(5):e25600) doi:10.2196/25600

XSL•FO
RenderX

KEYWORDS

COVID-19; health literacy; eHealth literacy; COVID-19–related health behavior questionnaire; Chinese college students

Introduction

COVID-19–Related Health Behaviors

The COVID-19 pandemic, a major public health emergency, has become a serious burden not only for China but also worldwide in terms of threatening people's lives as well as their mental health. At present, governments and institutions worldwide have introduced a series of policies to curb the spread of the epidemic, which have prevented and delayed the spread of COVID-19 to a certain extent [1]. However, vigilance continues to be required as global spread of COVID-19 is still very grim. World Health Organization data showed that as of March 22, 2021, the number of confirmed cases worldwide was 123,968,187, and 2,729,330 people have died worldwide [2]. Fortunately, new coronavirus vaccines have been developed. However, as current supply is insufficient, with potentially inconsistent quality among the different vaccines and manufacturers, many people have expressed suspicion and fear regarding vaccines [3] . Therefore, public preventive measures remain key to slowing the spread of COVID-19.

As a major public health event, the COVID-19 pandemic has required action from both governments and individuals. Since the COVID-19 outbreak, the Chinese government has implemented family-based isolation measures and other COVID-19 response measures, such as disease prevention strategies and advice for psychological adjustment [4]. In January 2020, the Guidelines for Public Protection Against Pneumonia Caused by the Novel Coronavirus Infection, which were compiled by the Chinese Center for Disease Control and Prevention (China CDC) [5], were provided to the public for free. These guidelines, which contain COVID-19 knowledge and prevention advice, such as personal protection, home medical observation, and psychological counseling, have been widely disseminated on online platforms such as social media. In addition to government measures, personal health behaviors also play a vital role in curbing the spread of COVID-19. Many previous studies have shown that public compliance with preventive behaviors can help reduce the spread of COVID-19 [6-8]. In addition to preventive behaviors, people's conventional health behaviors, such as physical exercise and diet, also play an important role in maintaining physical and mental health during the COVID-19 epidemic. Therefore, it is important to understand the characteristics of COVID-19–related health behaviors. In this study, we define COVID-19–related health behaviors as those that are in response to the perceived threat of COVID-19, including government-recommended preventive behaviors and self-adaptive conventional health behaviors. It is noted that few studies have developed specific tools by which to assess the degree of public COVID-19–related health behavior participation. A previous study in the United States developed a 9-item scale based on health behavior recommendations proposed by the US Centers for Disease Control and Prevention to assess individual health behaviors [9]. However, due to the different national conditions of the United States and China, this tool may not be suitable for the Chinese population. Through a literature review, we found that most research on COVID-19 health behavior in China was conducted using a single item or a count of checklist items; the reliability and validity of such approaches remain unclear. Therefore, objectively measuring people's COVID-19–related health behaviors and conducting research into the factors related to such behaviors may help formulate effective interventions for COVID-19 prevention and control.

Health Literacy

Due to the implementation of family isolation, people have had to rely on internet searches to obtain relevant COVID-19 health information (eg, prevention information and control guidelines), which may be associated with their COVID-19–related health behaviors. However, early studies have shown that although the internet can provide resources for individuals seeking health information, people do not always correctly utilize these resources to solve health problems [10]. Additionally, previous studies have demonstrated that individuals with greater health literacy may more precisely distinguish the authenticity and accuracy of COVID-19–related information on media platforms and have superior understanding of health information in general [11,12].

Health literacy is defined as an individual's ability to acquire, process, and understand basic health information and services in order to make appropriate health decisions [13]. Such literacy has long been associated with indicators of health [14]. A previous study showed that health literacy is associated with depression and health-related quality of life during the COVID-19 pandemic [15]. Specifically, medical students with higher health literacy have been shown to exhibit less fear of COVID-19 [16]. Additionally, adolescents' health literacy was shown to be significantly related to handwashing-related knowledge and behavior during the pandemic's early stages in Norway [17]. These examples indirectly illustrate that individuals with adequate health literacy are more likely to adopt COVID-19–related health behaviors, and the health literacy skill framework suggests that individuals' health behaviors were affected by their health literacy skills [18]. Therefore, we hypothesized that health literacy would be positively correlated with COVID-19–related health behaviors.

eHealth Literacy

In contrast, eHealth literacy specifically should be emphasized because while not everyone has a high level of health literacy during the COVID-19 crisis, they can still search, acquire, and utilize health information through the internet, allowing them to adopt relevant health behaviors based on this information. Health and eHealth literacy have distinct but related definitions, as both emphasize the individual's ability to collect, evaluate, and utilize health information; however, unlike health literacy, eHealth literacy focuses on the ability to obtain and apply online health information via electronic media [10]. Therefore, the hypothesis concerning the relationship between health literacy and COVID-19–related health behaviors may be extended to

XSL•FO
RenderX

include eHealth literacy, even though specific associations may be different for the two types of literacy.

Although previous studies emphasized the importance of eHealth literacy during COVID-19 [19,20]—and in the research of netizens [21], health care workers [22], and nursing students [23] it was found that eHealth literacy was related to COVID-19 prevention behavior—few studies have focused on the relationship between eHealth literacy and other COVID-19–related health behaviors other than preventive behaviors, especially among college students. Conventional health behaviors, such as physical exercise and nutritious eating, are also COVID-19–related health behaviors. In addition, a literature review indicated that eHealth literacy, healthy behaviors, and health outcomes are related [24]. Finally, although a previous study found that college students with a low level of eHealth literacy engaged in poor health-promoting behaviors [25], this relationship has not been adequately explored in a pandemic environment. Therefore, it is necessary to further analyze the relationship between eHealth literacy and COVID-19–related health behaviors. Overall, this online cross-sectional survey in China had two purposes: (1) develop and verify a COVID-19–related health behavior questionnaire and explore its status and structure and (2) study the relationship between health literacy, eHealth literacy, and COVID-19–related health behaviors.

Methods

Study Participants

A snowball sampling method was used to recruit the subjects. First, we published a recruitment announcement on QQ, which is a social media platform commonly used by Chinese college students. After screening, we identified 20 college students as the first group of participants from 20 universities across five regions of eastern, western, southern, northern, and central China. Each region included four universities to ensure that the sample broadly represented college students from all of China's regions.

Then, these participants anonymously completed an internet-based, cross-sectional questionnaire and were asked to send the link to the blank questionnaire to their classmates. Participation was voluntary and no financial incentive was given for participating in the study. The purpose of both the survey and informed consent form was displayed on the first page of the online questionnaire. If they agreed to participate, subjects could click the *Next* button to complete the questionnaire; if they chose not to participate, they could click the *X* button. The questionnaire could be completed in 15 minutes, and there was only one chance for a given internet protocol address from which to complete the questionnaire.

This study was approved by the Ethics Committee of Xiangya School of Public Health of Central South University. The survey started on May 10, 2020, and ended on May 20, 2020, recruiting a total of 2152 participants. Of these, 279 were excluded for their illogical completion of the questionnaire (eg, they took less than 5 minutes to complete it or their age was under 10 years). Finally, 1873 valid questionnaires were considered for our data analysis.

Measurements

Health Literacy

Health literacy was measured using the National Health Literacy Survey Questionnaire, as compiled by the China Health Education Center [26]. The overall Cronbach α of the scale was .95, and the Spearman-Brown coefficient was .94, indicating strong psychometric properties with minor measurement invariance [27]. The questionnaire was divided into three aspects: (1) basic health knowledge and literacy, (2) healthy lifestyle and behavior literacy, and (3) basic health skills literacy. These aspects covered the following six types of health problem literacy: scientific health, infectious disease prevention, chronic disease prevention, safety and emergency, basic medical information, and health-related information. The questionnaire consisted of 50 questions, including 34 multiple-choice questions and 16 multiple-answer questions (Multimedia Appendix 1). The total score ranged from 0 to 66 points, with 1 point per multiple-choice question and 2 points per multiple-answer question. A higher total score indicated a higher health literacy level; specifically, a total score of 53 or more indicated that the individual had adequate health literacy [28]. The Cronbach α coefficient for the questionnaire in this study was .83.

eHealth Literacy

eHealth literacy was measured using the Chinese version [29] of the eHealth Literacy Scale for College Students, originally developed by Norman and Skinner [30] to assess an individual's abilities in terms of accessing, understanding, and evaluating health information from electronic media and utilizing this information to solve health problems. The scale has eight items in total; each item was rated on a scale of 1 (disagree) to 5 (agree) (Multimedia Appendix 1). The total score ranged from 8 to 40; the higher the score, the higher the eHealth literacy. The Cronbach α coefficient for the scale in this study was .92.

COVID-19–Related Health Behaviors

We designed a questionnaire for measuring COVID-19–related health behaviors. First, we conducted an extensive literature review and identified 2 categories—government-recommended preventive behaviors and self-adaptive conventional health behaviors—to serve as the questionnaire's dimensions and overall framework. Based on the literature and the Guidelines for Public Protection Against Pneumonia Caused by the Novel Coronavirus Infection [5], we designed 24 items, 12 in each of the two categories. Second, an expert team composed of emergency management experts, public health experts, psychologists, respiratory doctors, and nurses evaluated the 24 items. According to their recommendations, seven items were deleted from the self-adaptive conventional health behaviors category, while three items were added to the preventive behaviors dimension because the experts believed that behaviors to prevent infection were the most important to adopt during the COVID-19 pandemic. Third, the 20 remaining items were sent to a different expert group, consistent with the previous group's composition, for review. According to the group's

XSL•FO
RenderX

opinions, five items were deleted from the preventive behaviors category.

Finally, a 15-item questionnaire was formed (Multimedia Appendix 1). We recruited 85 college students (41 males [48%] and 44 females [52%] with mean age of 20.1 years, SD 1.4) to conduct an online preliminary assessment. A 5-point Likert scale was used to test whether individuals understood each item's description, and they were asked whether they thought the item needed to be modified. The results showed that all respondents understood the descriptions (mean 4.48, SD 0.30), as all respondents indicated that no changes were required.

The questionnaire's 15 items focused on problem solving, such as wearing masks and handwashing; seeking social support, such as contacting relatives and friends online; distraction, denial, or avoidance, such as smoking and drinking; and positive appraisals, such as physical exercise and maintaining a reasonable diet. Participants were asked to indicate their practice of each behavior over the past 2 weeks. Each item used a 5-level scoring method: 1 (none of the time), 2 (a small amount of the time), 3 (sometimes), 4 (most of the time), and 5 (almost all of the time). It should be noted that the scores for smoking and drinking were reversed. Therefore, for this questionnaire, the higher the score, the healthier and more positive a COVID-19–related health behavior was considered.

Covariates

Covariates included sociodemographic characteristics (ie, age, gender, residence, college year, academic major, and self-rated family economic level) and information related to COVID-19 (ie, self-reported health status, family members or friends infected with the coronavirus, and self-reported coronavirus knowledge level) (Multimedia Appendix 1). The self-reported COVID-19 knowledge level was evaluated via five questions, regarding the source of COVID-19 infection, incubation period, main transmission route, susceptible population, and primary clinical manifestations (Multimedia Appendix 1). The questions and answers were based on the COVID-19 Diagnosis and Treatment Protocol (Tentative Version Seven) that was issued by the National Health Commission of China [31]. For these questions, 1 point was assigned to correct answers and no points were assigned to answers that were either incorrect or unknown. The total score ranged from 0 to 5; the higher the score, the higher the coronavirus knowledge level.

Statistical Analysis

Descriptive analysis was used to summarize participants' sociodemographic information and other variables. Categorical variables were described using a frequency and percentage, whereas continuous variables were described by mean and SD. Then, to evaluate the factor structure and structural validity of the COVID-19–related health behavior questionnaire, a principal component exploratory factor analysis (EFA) with a varimax rotation and a confirmatory factor analysis (CFA) were used. It should be noted that we randomly divided the total samples into Sample A (511 men and 426 women), which was used for EFA, and Sample B (458 men and 478 women), which was used for CFA.

The Kaiser-Meyer-Olkin (KMO>0.50) measure of sampling adequacy and the Bartlett test of sphericity (*P* value of the Bartlett test <.05) were used to assess the questionnaire's suitability for factor analysis [32]. Eigenvalues of greater than 1 were used to determine the number of factors, and a factor loading of greater than 0.50 was regarded as the criterion for keeping items.

Then, CFA, using SPSS Amos, version 23.0 (IBM Corp), was conducted to examine the factor structure of the COVID-19–related health behavior questionnaire. The CFA's goodness of fit was evaluated using the following indicators: goodness-of-fit index (GFI >0.90), adjusted GFI (AGFI >0.90), root mean square error of approximation (RMSEA <0.80), comparative fit index (CFI >0.90), Tucker-Lewis index (TLI >0.90), and standardized root mean square residual (SRMR <0.80) [33]. The Cronbach α coefficient was used to evaluate the questionnaire's internal consistency reliability.

The Student *t* test, or one-way analysis of variance (ANOVA), was used to evaluate the statistical differences in the distribution of COVID-19–related health behaviors across different sociodemographic characteristics and health literacy levels. Cohen *d* was used to evaluate effect size. Pearson correlations were used to quantify the bivariate associations between self-reported coronavirus knowledge level, health literacy, eHealth literacy, and COVID-19–related health behaviors. Hierarchical linear regression was used to analyze the relationship between health literacy, eHealth literacy, and COVID-19–related health behaviors. Correlation magnitudes were reported as standardized regression coefficients (β). SPSS, version 23.0 (IBM Corp), was used to perform statistical analyses. All statistical significance levels were set at α=.05, and the statistical tests were two-tailed.

Results

Participant Characteristics

The mean age of the participants was 19.6 years (SD 1.8), ranging from 18 to 25 years. Of the 1873 participants, 1505 (80.4%) had good self-reported health status, while 66 (3.5%) had family members or friends infected with coronavirus. The mean self-reported coronavirus knowledge level score was 4.11 (SD 1.18), and nearly half of the participants (903/1873, 48.2%) correctly answered five COVID-19 knowledge questions. The mean health literacy score was 50.05 (SD 9.55), and 781 out of 1873 participants (41.7%) had adequate health literacy. The mean eHealth literacy score was 30.16 (SD 6.31). For further details, see Table 1.

XSL•FO
RenderX

Table 1. Descriptive statistics of sociodemographic characteristics of the study participants.

Characteristic	Participants (N=1873), n (%)
Gender	
Male	969 (51.7)
Female	904 (48.3)
Residence	
Urban	871 (46.5)
Rural	1002 (53.5)
College year	
Freshman	394 (21.0)
Sophomore	594 (31.7)
Junior	507 (27.1)
Senior	378 (20.2)
Academic major	
Medicine	851 (45.4)
Others	1022 (54.6)
Family economic level	
High	568 (30.3)
Medium	1114 (59.5)
Low	191 (10.2)
Self-reported health status	
Good	1505 (80.4)
Medium	336 (17.9)
Bad	32 (1.7)
Family member or friend infected with coronavirus	
Yes	66 (3.5)
No	1807 (96.5)
Self-reported coronavirus knowledge level	
0 correct responses	34 (1.8)
1 correct response	69 (3.7)
2 correct responses	96 (5.1)
3 correct responses	156 (8.3)
4 correct responses	615 (32.8)
5 correct responses	903 (48.2)
Health literacy level	
Inadequate	1092 (58.3)
Adequate	781 (41.7)

Internal Consistency and Structural Validity of the COVID-19–Related Health Behavior Questionnaire

First, we used the total sample for a presupposed two-factor structure analysis: 10 items for government-recommended preventive behaviors and five items for self-adaptive conventional health behaviors. However, the two-factor model for a 15-item questionnaire did not satisfactorily fit the data in the total sample (GFI=0.804; AGFI=0.735; RMSEA=0.131; CFI=0.658; TLI=0.597; SRMR=0.127). Therefore, we performed an EFA and a CFA. Table 2 shows the EFA results for each item of the COVID-19–related health behavior questionnaire. The value of the KMO test was 0.85, and the *P* value of Bartlett test was <.001. For each item, the factor loading ranged from 0.50 to 0.75 (all ≥0.50), such that the conditions for conducting EFA were satisfied. The results revealed two

XSL•FO
RenderX

factors with eigenvalues greater than 1, accounting for 45.96% of the total variance. Factor 1 was comprised of eight items; its eigenvalue was 4.40 and it explained 29.34% of the variance. Factor 2 was comprised of seven items; its eigenvalue was 2.49 and it explained 16.62% of the variance.

Next, we performed CFA with Sample B. The results showed that a two-factor model had adequate goodness of fit (GFI=0.948; AGFI=0.919; RMSEA=0.068; CFI=0.924; TLI=0.909; SRMR=0.057). Based on the items that loaded on each factor, we preliminarily named the two factors COVID-19–specific precautionary behaviors, which contained eight items related to prevention, and conventional health behaviors, which contained seven items related to conventional health behaviors. Among the items, the highest score was for *cover your mouth and nose when you cough or sneeze* and the lowest score was for *insist on physical exercise*. For the 15 items in the total sample, the Cronbach α internal reliability coefficient was .79.

Table 2. Exploratory factor analysis and descriptive statistics of the COVID-19–related health behavior questionnaire for college students.

Item	Sample A factor loading (n=937)		Total sample score[a] (N=1873), mean (SD)
	Factor 1	Factor 2	
Maintain hand hygiene	0.75	0.20	3.84 (0.96)
Open windows for ventilation to maintain air circulation	0.70	0.16	3.87 (0.96)
Wear a mask when going out	0.69	0.06	4.10 (0.97)
Reduce instances of going to public places	0.63	0.30	3.70 (0.95)
Cover your mouth and nose when you cough or sneeze	0.61	0.19	4.28 (0.96)
Disinfect daily necessities	0.60	0.36	3.52 (1.03)
Contact relatives and friends online	0.52	0.30	3.50 (0.98)
Communicate or confide with others when you are in trouble	0.51	0.22	3.54 (1.07)
Maintain adequate nutrition and balanced diet	0.23	0.66	2.86 (1.06)
Drink alcohol because of COVID-19	0.26	0.65	4.11 (1.16)
Insist on physical exercise	0.08	0.64	2.72 (1.21)
Smoke because of COVID-19	0.35	0.63	4.18 (1.28)
Follow the latest developments on COVID-19	0.30	0.63	3.07 (1.07)
Take body temperature frequently	0.25	0.53	3.24 (1.11)
Guarantee good sleep	0.34	0.50	3.25 (1.09)

[a]Responses were based on a 5-level scoring method: 1 (none of the time), 2 (a small amount of the time), 3 (sometimes), 4 (most of the time), and 5 (almost all of the time).

The Distribution of COVID-19–Related Health Behaviors Across Different Characteristics

The mean scores of COVID-19–related health behaviors, COVID-19–specific precautionary behaviors, and conventional health behaviors were 53.77 (SD 8.03), 30.36 (SD 5.17), and 23.41 (SD 3.87), respectively. The results of *t* tests and ANOVAs showed that the differences in COVID-19–related health behavior scores and COVID-19–specific precautionary behavior scores among college students with different genders, residences, college years, academic majors, family economic levels, self-reported health statuses, family members or friends infected with coronavirus, and health literacy levels were statistically significant ($P<.05$). In addition, there were also significant differences in conventional health behavior scores among the college students with different residences, college years, family economic levels, self-reported health statuses, family members or friends infected with coronavirus, and health literacy levels ($P<.05$). Further details are displayed in Multimedia Appendix 2.

Associations Between Self-Reported Coronavirus Knowledge Level, Health Literacy, eHealth Literacy, and COVID-19–Related Health Behaviors

Table 3 shows that self-reported coronavirus knowledge levels were positively associated with health literacy ($r=0.162$; $P<.001$), eHealth literacy ($r=0.220$; $P<.001$), and COVID-19–related health behaviors ($r=0.244$; $P<.001$). Additionally, health literacy was positively associated with eHealth literacy ($r=0.270$; $P<.001$) and COVID-19–related health behaviors ($r=0.338$; $P<.001$), while eHealth literacy was positively associated with COVID-19–related health behaviors ($r=0.476$; $P<.001$).

XSL•FO
RenderX

Table 3. Correlation matrix (Pearson *r* and two-tailed *P* value[a]) of self-reported coronavirus knowledge level, health literacy, eHealth literacy, and health and precautionary behaviors.

Variable	Self-reported coronavirus knowledge level	Health literacy	eHealth literacy	COVID-19–related health behaviors	COVID-19–specific precautionary behaviors	Conventional health behaviors
Self-reported coronavirus knowledge level						
r	1	0.162	0.220	0.244	0.283	0.129
P value	—[b]	<.001	<.001	<.001	<.001	<.001
Health literacy						
r	0.162	1	0.270	0.338	0.288	0.316
P value	<.001	—	<.001	<.001	<.001	<.001
eHealth literacy						
r	0.220	0.270	1	0.476	0.471	0.357
P value	<.001	<.001	—	<.001	<.001	<.001
COVID-19–related health behaviors						
r	0.244	0.338	0.476	1	0.918	0.848
P value	<.001	<.001	<.001	—	<.001	<.001
COVID-19–specific precautionary behaviors						
r	0.283	0.288	0.471	0.918	1	0.568
P value	<.001	<.001	<.001	<.001	—	<.001
Conventional health behaviors						
r	0.129	0.316	0.357	0.848	0.568	1
P value	<.001	<.001	<.001	<.001	<.001	—

[a]For all associations, the correlation is significant at a significance level of .05 (two-tailed).

[b]Not applicable.

Factors Associated With COVID-19–Related Health Behaviors: Multivariable Analyses

Hierarchical linear regression analysis was conducted with the score of the COVID-19–related health behaviors (COVID-19–specific precautionary behaviors and conventional health behaviors) as the dependent variable, while the independent variables consisted of sociodemographic characteristics, health literacy, and eHealth literacy. In Block 1, the following sociodemographic characteristics were entered first, accounting for 17.4% of the variance in COVID-19–specific precautionary behaviors (R^2=0.174; $F_{13, 1859}$=33.890; P<.001) and 7.4% of the variance in conventional health behaviors (R^2=0.074; $F_{13, 1859}$=13.393; P<.001): gender, residence, college year, academic major, family economic level, self-reported health status, having a family member or friend infected with coronavirus, and self-reported coronavirus knowledge level. The score for health literacy was entered into Block 2 and contributed to explaining an additional 4.2% of the variance in COVID-19–specific precautionary behaviors (R^2=0.216; $F_{14, 1858}$=40.692; P<.001) and an additional 6.4% of the variance in conventional health behaviors (R^2=0.138; $F_{14, 1858}$=24.045; P<.001). Finally, the score for eHealth literacy was entered into Block 3 and contributed to explaining an additional 11.5% of the variance in COVID-19–specific precautionary behaviors (R^2=0.332; $F_{15, 1857}$=67.449; P<.001) and an additional 6.5% of the variance in conventional health behaviors (R^2=0.203; $F_{14, 1858}$=35.083; P<.001). The linear regression results are presented in Multimedia Appendix 3.

Multimedia Appendix 3 illustrates that the listed sociodemographic characteristics, other than residence and academic major, were related to COVID-19–related health behaviors. Overall, participants were more likely to have superior COVID-19–specific precautionary behavior scores if they were female versus male (β=.153; P<.001), freshmen versus seniors (β=.094; P<.001), juniors versus seniors (β=–.061; P=.02), from a family with a medium versus low economic level (β=.122; P<.001), had a self-reported health status that was good versus bad (β=.205; P=.001), had a family member or friend infected with coronavirus versus not infected (β=.095; P<.001), and had a higher self-reported coronavirus knowledge level (β=.156; P<.001). Conventional health behaviors were more likely to be performed by individuals who were sophomores versus seniors (β=.074; P<.001), from a family with a high or medium economic level (high vs low: β=.082, P=.03; medium vs low: β=.077, P=.03), had good or medium self-reported health status (good vs bad: β=.273, P<.001; medium vs bad: β=.171, P=.008).

After adjusting for covariates, both health literacy and eHealth literacy were positively associated with COVID-19–specific

XSL•FO
RenderX

precautionary behaviors ($\beta_{\text{health literacy}}$=.149, $\beta_{\text{eHealth literacy}}$=.368; *P*<.001) and conventional health behaviors ($\beta_{\text{health literacy}}$=.219, $\beta_{\text{eHealth literacy}}$=.277; *P*<.001). Among the related factors, eHealth literacy was most strongly associated with COVID-19–related health behaviors.

Discussion

Principal Findings

Based on the China CDC's guidelines and expert advice, we developed and verified a COVID-19–related health behavior questionnaire among college students. The questionnaire was found to be valid and reliable for assessing health behaviors during the pandemic. This study also examined whether health literacy and eHealth literacy were associated with the COVID-19–related health behaviors of college students. To the best of our knowledge, this is one of few studies to consider the relationship between these factors for this population group. The study results showed that health and eHealth literacy were positively related to COVID-19–related health behaviors among college students.

The factor analysis showed that the COVID-19–related health behavior questionnaire had a two-factor structure, which is consistent with the two-factor structure we had previously assumed. That is, the eight items for Factor 1 primarily reflected individual protective measures, such as wearing a mask, maintaining hand hygiene, and reducing the amount of time that one leaves the house. Therefore, Factor 1 was named *COVID-19–specific precautionary behaviors*. Furthermore, the seven items for Factor 2 reflected individuals' conventional health behaviors, such as physical exercise and diet during the COVID-19 pandemic; therefore, it was named *conventional health behaviors*. In addition, the results showed that the internal consistency reliability of the questionnaire was adequate. Based on these two relatively simple psychometric indicators, the COVID-19–related health behavior questionnaire appears to have satisfactory internal consistency reliability and construct validity in this context, so that it could be used among Chinese college students. This questionnaire may be useful for government officials and school health educators to assess the COVID-19–related health behaviors of college students in order to carry out targeted interventions.

The results of the descriptive analysis showed that college students' COVID-19–specific precautionary behaviors were high, and the scores of eight items were all higher than 3.5, among which the score of *cover your mouth and nose when you cough or sneeze* was the highest. A European study of adolescents obtained similar results, namely that adolescents seemed to be generally aware of the recommendations regarding protective behaviors during the COVID-19 crisis [17]. In contrast, this study found that college students' conventional health behaviors were poor, especially in terms of their scores for physical exercise and maintaining both adequate nutrition and a balanced diet, which were lower than the median (3 points). Zhao et al also found that after the COVID-19 outbreak, only a small number of Chinese people participated in physical exercise, and most consumed vegetables and fruits even less frequently than usual each week [34]. This may be related to affective disorders, such as anxiety and depression during the COVID-19 pandemic. A cross-sectional study suggested that less physical exercise was found to negatively impact anxiety and depressive symptoms in college students during the COVID-19 pandemic [35]. However, in turn, anxiety and depression can cause a lack of motivation to exercise and eat healthily. Thus, this study indirectly reflects the changes that the COVID-19 pandemic caused in people's lifestyles and behaviors. Additionally, previous studies showed that unhealthy behaviors were related to COVID-19 mortality rates and hospital admissions, and although they provided some specific lifestyle recommendations [36,37], few studies have analyzed intervenable factors related to COVID-19–related health behaviors during the pandemic.

In this study, after adjusting for sociodemographic characteristics, health and eHealth literacy were shown to be significantly related to COVID-19–related health behaviors, which supports the findings of previous research on the relationship between health literacy and COVID-19–related attitudes and behaviors [38]. For example, a study in Hong Kong demonstrated that health literacy significantly and positively correlated with hand hygiene habits in older persons [39], consistent with a study of adolescents [17], indicating that health literacy is closely related to COVID-19 health protection behaviors. In addition, health literacy's positive relationship with physical exercise, diet, and sleep has been confirmed by numerous studies [40-42]; likewise, this study confirmed this relationship. This indicates that, even during the COVID-19 pandemic, health literacy remains related to healthy behaviors. Nevertheless, these results cannot explain the causal relationship between health literacy and COVID-19–related health behaviors. However, based on previous research experience [43], individuals with a high level of health literacy are more likely to adopt positive behaviors in their response to the COVID-19 pandemic. Additionally, although it was obviously difficult to improve the health literacy of individuals on such a short-term basis, significantly curbing the disease's spread will involve finding new intervention strategies and incorporating the abundance of information related to COVID-19 on the internet.

This study showed that eHealth literacy was also positively correlated with college students' COVID-19–related health behaviors. Similarly, a previous study illustrated that people who had obtained more health information online related to COVID-19 were more frequently involved in various types of preventive behaviors [44]. Researchers have attributed this association to risk perception; COVID-19 pandemic information on the internet exacerbates concerns and may prompt users to adopt protective behaviors [45]. Further, a related study of college students reported that their eHealth literacy is associated with regular exercise, healthy eating, and routine sleep [46]. Moreover, a study found that eHealth literacy is related to disease prevention behaviors, such as finding vaccination information [47]. These results indirectly support the association between eHealth literacy and COVID-19–related health behaviors. Furthermore, a recent study showed that eHealth literacy is related to preventive behaviors in relation to the COVID-19 pandemic [21], consistent with findings in this study.

XSL•FO
RenderX

However, the previous study had notably few people under the age of 20 years, so that this study, which included a substantial number of subjects under this age, further expanded upon their findings.

This study also found that a higher level of eHealth literacy was associated with superior conventional health behaviors, indicating that college students with a higher level of eHealth literacy in China could maintain healthy lifestyles during the COVID-19 pandemic. While previous studies have paid more attention to health protection behaviors, few have examined other behaviors during the COVID-19 pandemic, such as exercise, diet, sleep, smoking, and consuming alcohol. For example, while the significance of COVID-19–specific precautionary behaviors focuses on reducing the possibility of an individual becoming infected with the coronavirus, conventional health behaviors provide for greater reflection in an individual's daily life during this period of isolation from family due to COVID-19 prevention and control measures. Conventional health behaviors emphasize the ability to adapt to life events, which is particularly important when promoting physical and mental health, such as maintaining resilience, reducing and avoiding the occurrence of other diseases, and relieving both mental pressures and stress. Therefore, this study suggests that improving health and eHealth literacy may help college students adopt healthy lifestyles and adapt to the isolation that has been caused by the COVID-19 pandemic.

Compared to health literacy, eHealth literacy was a stronger predictor of COVID-19–related health behaviors, which may be due to COVID-19 being a sudden infectious disease. That is, the nature of this disease causes people to lack sufficient understanding about it, which leads to a lack of knowledge about how to take preventive measures. Since the Chinese government implemented family isolation measures during the COVID-19 crisis, they published knowledge and guidelines on the internet about various COVID-19 prevention and control measures. A previous study showed that individuals with higher levels of eHealth literacy more often search for health information on the internet [48], which means that college students with a high level of eHealth literacy may acquire more COVID-19 information online, enabling them to adopt health behaviors. Considering that health literacy is the result of personal, long-term learning and behavioral practice [49], it is difficult to change this feature in a short-term period. Therefore, this study suggests that it is particularly important to pay more attention to improving individual eHealth literacy.

This study's results have important value in terms of public health, especially concerning the prevention and control of COVID-19 in universities. Many universities around the world have resumed studies on campus, causing numerous students to gather on small campuses, which places greater pressure on the need for COVID-19 prevention and control. For instance, once an asymptomatic patient who is infected with the coronavirus appears at school, they could transmit the infection to other students, especially persons not practicing health protection measures. Since college students are the primary users of electronic media and the internet in general, the feasibility and practical effects of related prevention and control interventions through eHealth literacy are potentially significant.

Regarding limitations, snowball sampling is nonprobability sampling that is generally used when it is difficult to identify the members of the population. Since the Chinese government adopted the family segregation policy when we conducted the survey and schools did not open, it was difficult for us to use probability sampling when conducting the survey. Snowball sampling has the possibility of sampling and selection bias, which limits the extrapolation of our research. Second, a cross-sectional survey cannot explain the causal relationships among health literacy, eHealth literacy, and COVID-19–related health behaviors. Third, in the survey, we did not use a standardized questionnaire to assess participants' knowledge of COVID-19, which may not adequately reflect knowledge of COVID-19. Recent research developed a health literacy scale related to the coronavirus [50]; future research could utilize this scale to obtain more in-depth understanding. Finally, the COVID-19–related health behavior questionnaire was compiled according to China's cultural background as well as the state's COVID-19 prevention and control guidelines. Thus, it may not be applicable to other countries, which limits the inferences one may make from these results.

Conclusions

The COVID-19–related health behavior questionnaire was a valid and reliable measure for assessing health behaviors during the pandemic. The results of this study suggest that college students with higher health literacy and eHealth literacy can more actively adopt COVID-19–related health behaviors. Additionally, compared to health literacy, eHealth literacy is more closely related to COVID-19–related health behaviors. These findings are significant for COVID-19 prevention and control in the college environment, including for relevant health education activities or other intervention measures based on health and eHealth literacy that can be carried out in a timely manner to enhance COVID-19–related health behaviors and reduce the risk of infection among college students.

Acknowledgments

We would like to thank all participants who participated in this study.

Authors' Contributions

SL and GC designed the study, conducted surveys and data analysis, and wrote the first draft of the manuscript. ACK and SC revised all versions of the manuscript. HX made a number of constructive suggestions and revised the final draft of the manuscript. All authors have read and agreed to the final draft of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaires.
[PDF File (Adobe PDF File), 588 KB - jmir_v23i5e25600_app1.pdf]

Multimedia Appendix 2
Comparison of COVID-19–related health behaviors among college students with different demographic characteristics.
[PDF File (Adobe PDF File), 214 KB - jmir_v23i5e25600_app2.pdf]

Multimedia Appendix 3
Factors associated with COVID-19–related health behaviors in Chinese college students.
[PDF File (Adobe PDF File), 290 KB - jmir_v23i5e25600_app3.pdf]

References

1. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020 Aug;584(7820):262-267. [doi: 10.1038/s41586-020-2404-8] [Medline: 32512578]
2. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2021-03-22]
3. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: Addressing vaccine hesitancy and fostering vaccine confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
4. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 Apr 10;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
5. Chinese Center for Disease Control and Prevention. Guidelines for Public Protection Against Pneumonia Caused by New Coronavirus Infection. Beijing, China: Chinese People's Medical Publishing House; 2020.
6. Nguyen NPT, Hoang TD, Tran VT, Vu CT, Siewe Fodjo JN, Colebunders R, et al. Preventive behavior of Vietnamese people in response to the COVID-19 pandemic. PLoS One 2020 Sep 9;15(9):e0238830 [FREE Full text] [doi: 10.1371/journal.pone.0238830] [Medline: 32903286]
7. Chen X, Chen H. Differences in preventive behaviors of COVID-19 between urban and rural residents: Lessons learned from a cross-sectional study in China. Int J Environ Res Public Health 2020 Jun 20;17(12):4437 [FREE Full text] [doi: 10.3390/ijerph17124437] [Medline: 32575700]
8. Duan T, Jiang H, Deng X, Zhang Q, Wang F. Government intervention, risk perception, and the adoption of protective action recommendations: Evidence from the COVID-19 prevention and control experience of China. Int J Environ Res Public Health 2020 May 13;17(10):3387 [FREE Full text] [doi: 10.3390/ijerph17103387] [Medline: 32414013]
9. Toussaint LL, Cheadle AD, Fox J, Williams DR. Clean and contain: Initial development of a measure of infection prevention behaviors during the COVID-19 pandemic. Ann Behav Med 2020 Sep 1;54(9):619-625 [FREE Full text] [doi: 10.1093/abm/kaaa064] [Medline: 32856691]
10. 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]
11. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
12. Baines D, Elliott R. Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers 20-06. IDEAS. Birmingham, UK: Department of Economics, University of Birmingham; 2020. URL: https://ideas.repec.org/p/bir/birmec/20-06.html [accessed 2020-11-20]
13. Berkman ND, Davis TC, McCormack L. Health literacy: What is it? J Health Commun 2010;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
14. Greenhalgh T. Health literacy: Towards system level solutions. BMJ 2015 Feb 24;350:h1026. [Medline: 25712067]
15. Nguyen HC, Nguyen MH, Do BN, Tran CQ, Nguyen TTP, Pham KM, et al. People with suspected COVID-19 symptoms were more likely depressed and had lower health-related quality of life: The potential benefit of health literacy. J Clin Med 2020 Mar 31;9(4):965 [FREE Full text] [doi: 10.3390/jcm9040965] [Medline: 32244415]
16. Nguyen HT, Do BN, Pham KM, Kim GB, Dam HTB, Nguyen TT, et al. Fear of COVID-19 Scale-Associations of its scores with health literacy and health-related behaviors among medical students. Int J Environ Res Public Health 2020 Jun 11;17(11):4164 [FREE Full text] [doi: 10.3390/ijerph17114164] [Medline: 32545240]

XSL•FO
RenderX

17. Riiser K, Helseth S, Haraldstad K, Torbjørnsen A, Richardsen KR. Adolescents' health literacy, health protective measures, and health-related quality of life during the Covid-19 pandemic. PLoS One 2020 Aug;15(8):e0238161 [FREE Full text] [doi: 10.1371/journal.pone.0238161] [Medline: 32857806]
18. Squiers L, Peinado S, Berkman N, Boudewyns V, McCormack L. The health literacy skills framework. J Health Commun 2012 Oct;17 Suppl 3:30-54. [doi: 10.1080/10810730.2012.713442] [Medline: 23030560]
19. 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]
20. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
21. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: Cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
22. Do BN, Tran TV, Phan DT, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health literacy, eHealth literacy, adherence to infection prevention and control procedures, lifestyle changes, and suspected COVID-19 symptoms among health care workers during lockdown: Online survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
23. Yuan T, Liu H, Li XD, Liu HR. Factors affecting infection control behaviors to prevent COVID-19: An online survey of nursing students in Anhui, China in March and April 2020. Med Sci Monit 2020 Oct 11;26:e925877 [FREE Full text] [doi: 10.12659/MSM.925877] [Medline: 33040074]
24. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
25. Yang S, Luo Y, Chiang C. The associations among individual factors, eHealth literacy, and health-promoting lifestyles among college students. J Med Internet Res 2017 Jan 10;19(1):e15 [FREE Full text] [doi: 10.2196/jmir.5964] [Medline: 28073739]
26. Li Y. Introduction of 2012 Chinese residents health literacy monitoring program. Chin J Health Educ 2014;30(6):563-565. [doi: 10.16168/j.cnki.issn.1002-9982.2014.06.003]
27. Shen M, Hu M, Liu S, Chang Y, Sun Z. Assessment of the Chinese Resident Health Literacy Scale in a population-based sample in South China. BMC Public Health 2015 Jul 11;15:637 [FREE Full text] [doi: 10.1186/s12889-015-1958-0] [Medline: 26163383]
28. Wang W, Zhang Y, Lin B, Mei Y, Ping Z, Zhang Z. The urban-rural disparity in the status and risk factors of health literacy: A cross-sectional survey in central China. Int J Environ Res Public Health 2020 May 29;17(11):3848 [FREE Full text] [doi: 10.3390/ijerph17113848] [Medline: 32485790]
29. Guo S, Yu X, Sun Y, Nie D, Li X, Wang L. Adaptation and evaluation of Chinese version of eHEALS and its usage among senior high school students. Chin J Health Educ 2013;29(2):106-108. [doi: 10.16168/j.cnki.issn.1002-9982.2013.02.019]
30. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
31. COVID-19 diagnosis and treatment protocol (tentative version 7). National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/xcs/zhengcwj/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml [accessed 2020-11-20]
32. Williams B, Onsman A, Brown T. Exploratory factor analysis: A five-step guide for novices. Australas J Paramedicine 2010 Aug 02;8(3):1-13. [doi: 10.33151/ajp.8.3.93]
33. Carrión G, Nitzl C, Roldán J. Mediation analyses in partial least squares structural equation modeling: Guidelines and empirical examples. In: Latan H, Noonan R, editors. Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications. Cham, Switzerland: Springer; 2017:173-195.
34. Hu Z, Lin X, Chiwanda Kaminga A, Xu H. Impact of the COVID-19 epidemic on lifestyle behaviors and their association with subjective well-being among the general population in mainland China: Cross-sectional study. J Med Internet Res 2020 Aug 25;22(8):e21176 [FREE Full text] [doi: 10.2196/21176] [Medline: 32759103]
35. 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]
36. Hamer M, Kivimäki M, Gale CR, Batty GD. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. Brain Behav Immun 2020 Jul;87:184-187 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.059] [Medline: 32454138]
37. Smirmaul BPC, Chamon RF, de Moraes FM, Rozin G, Moreira ASB, de Almeida R, et al. Lifestyle medicine during (and after) the COVID-19 pandemic. Am J Lifestyle Med 2021 Aug;15(1):60-67. [doi: 10.1177/1559827620950276] [Medline: 33456421]
38. Wolf MS, Serper M, Opsasnick L, O'Conor RM, Curtis LM, Benavente JY, et al. Awareness, attitudes, and actions related to COVID-19 among adults with chronic conditions at the onset of the US outbreak: A cross-sectional survey. Ann Intern Med 2020 Apr 09;173(2):100-109 [FREE Full text] [doi: 10.7326/M20-1239] [Medline: 32271861]

XSL•FO
RenderX

39. Or PP, Wong BY, Chung JW. To investigate the association between the health literacy and hand hygiene practices of the older adults to help them fight against infectious diseases in Hong Kong. Am J Infect Control 2020 May;48(5):485-489 [FREE Full text] [doi: 10.1016/j.ajic.2019.12.021] [Medline: 32037202]
40. Klinker CD, Aaby A, Ringgaard LW, Hjort AV, Hawkins M, Maindal HT. Health literacy is associated with health behaviors in students from vocational education and training schools: A Danish population-based survey. Int J Environ Res Public Health 2020 Jan 20;17(2):671 [FREE Full text] [doi: 10.3390/ijerph17020671] [Medline: 31968667]
41. Barsell DJ, Everhart RS, Miadich SA, Trujillo MA. Examining health behaviors, health literacy, and self-efficacy in college students with chronic conditions. Am J Health Educ 2018 Jul 18;49(5):305-311. [doi: 10.1080/19325037.2018.1486758]
42. Luo H, Chen Z, Bell R, Rafferty AP, Gaskins Little NR, Winterbauer N. Health literacy and health behaviors among adults with prediabetes, 2016 Behavioral Risk Factor Surveillance System. Public Health Rep 2020;135(4):492-500. [doi: 10.1177/0033354920927848] [Medline: 32511939]
43. Zhang L, Seale H, Wu S, Yang P, Zheng Y, Ma C, et al. Post-pandemic assessment of public knowledge, behavior, and skill on influenza prevention among the general population of Beijing, China. Int J Infect Dis 2014 Jul;24:1-5 [FREE Full text] [doi: 10.1016/j.ijid.2014.01.003] [Medline: 24735679]
44. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: Cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
45. Sun Z, Yang B, Zhang R, Cheng X. Influencing factors of understanding COVID-19 risks and coping behaviors among the elderly population. Int J Environ Res Public Health 2020 Aug 13;17(16):5889 [FREE Full text] [doi: 10.3390/ijerph17165889] [Medline: 32823740]
46. Tsukahara S, Yamaguchi S, Igarashi F, Uruma R, Ikuina N, Iwakura K, et al. Association of eHealth literacy with lifestyle behaviors in university students: Questionnaire-based cross-sectional study. J Med Internet Res 2020 Jun 24;22(6):e18155 [FREE Full text] [doi: 10.2196/18155] [Medline: 32579126]
47. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth literacy and health behaviors affecting modern college students: A pilot study of issues identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]
48. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: A mediation analysis. J Med Internet Res 2017 May 16;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]
49. Edwards M, Wood F, Davies M, Edwards A. The development of health literacy in patients with a long-term health condition: The health literacy pathway model. BMC Public Health 2012;12:130 [FREE Full text] [doi: 10.1186/1471-2458-12-130] [Medline: 22332990]
50. Okan O, Bollweg TM, Berens E, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-related health literacy: A cross-sectional study in adults during the COVID-19 infodemic in Germany. Int J Environ Res Public Health 2020 Jul 30;17(15):5503 [FREE Full text] [doi: 10.3390/ijerph17155503] [Medline: 32751484]

Abbreviations

AGFI: adjusted goodness-of-fit index
ANOVA: analysis of variance
CFA: confirmatory factor analysis
CFI: comparative fit index
China CDC: Chinese Center for Disease Control and Prevention
EFA: exploratory factor analysis
GFI: goodness-of-fit index
KMO: Kaiser-Meyer-Olkin
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

XSL•FO
RenderX

Edited by G Eysenbach, C Basch; submitted 08.11.20; peer-reviewed by X Li, E Neter, K Dadaczynski; comments to author 23.12.20; revised version received 29.01.21; accepted 03.04.21; published 06.05.21.

Please cite as:
Li S, Cui G, Kaminga AC, Cheng S, Xu H
Associations Between Health Literacy, eHealth Literacy, and COVID-19–Related Health Behaviors Among Chinese College Students: Cross-sectional Online Study
J Med Internet Res 2021;23(5):e25600
URL: https://www.jmir.org/2021/5/e25600
doi:10.2196/25600
PMID:33822734

©Shaojie Li, Guanghui Cui, Atipatsa Chiwanda Kaminga, Sixiang Cheng, Huilan Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2021. 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 Behavioral Intentions Toward COVID-19 Vaccines: Theory-Based Content Analysis of Tweets

Siru Liu[1], BS; Jialin Liu[2], MD

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

[2]Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Jialin Liu, MD
Department of Medical Informatics
West China Hospital
Sichuan University
No 37 Wainan Guoxuexiang Street
Chengdu, 610041
China
Phone: 86 28 85422306
Email: dljl8@163.com

Abstract

Background: Acceptance rates of COVID-19 vaccines have still not reached the required threshold to achieve herd immunity. Understanding why some people are willing to be vaccinated and others are not is a critical step to develop efficient implementation strategies to promote COVID-19 vaccines.

Objective: We conducted a theory-based content analysis based on the capability, opportunity, motivation–behavior (COM-B) model to characterize the factors influencing behavioral intentions toward COVID-19 vaccines mentioned on the Twitter platform.

Methods: We collected tweets posted in English from November 1-22, 2020, using a combination of relevant keywords and hashtags. After excluding retweets, we randomly selected 5000 tweets for manual coding and content analysis. We performed a content analysis informed by the adapted COM-B model.

Results: Of the 5000 COVID-19 vaccine–related tweets that were coded, 4796 (95.9%) were posted by unique users. A total of 97 tweets carried positive behavioral intent, while 182 tweets contained negative behavioral intent. Of these, 28 tweets were mapped to capability factors, 155 tweets were related to motivation, 23 tweets were related to opportunities, and 74 tweets did not contain any useful information about the reasons for their behavioral intentions (κ=0.73). Some tweets mentioned two or more constructs at the same time. Tweets that were mapped to capability (P<.001), motivation (P<.001), and opportunity (P=.03) factors were more likely to indicate negative behavioral intentions.

Conclusions: Most behavioral intentions regarding COVID-19 vaccines were related to the motivation construct. The themes identified in this study could be used to inform theory-based and evidence-based interventions to improve acceptance of COVID-19 vaccines.

(J Med Internet Res 2021;23(5):e28118) doi:10.2196/28118

KEYWORDS

vaccine; COVID-19; behavior; tweet; intention; content analysis; Twitter; social media; acceptance; threshold; willing; theory; model; infodemiology; infoveillance

Introduction

The global COVID-19 pandemic is affecting 219 countries worldwide [1]. An important component of managing COVID-19 is preventing the infection [2,3]. Fortunately, development of vaccines against the disease is progressing well. In December 2020, the US Food and Drug Administration authorized COVID-19 vaccines for emergency use. Another currently pressing issue is how to increase vaccine acceptance rates [4]. In previously published survey studies, the acceptance rate of COVID-19 vaccines was a concern. Of 672 participants in the United States, approximately 67% said they would be willing to receive a vaccine [5]. It is necessary to vaccinate an estimated 55%-82% of the population to create herd immunity

XSL•FO
RenderX

and slow the spread of a pandemic [6]. Therefore, it is critical to understand why some people are willing to be vaccinated and others are not.

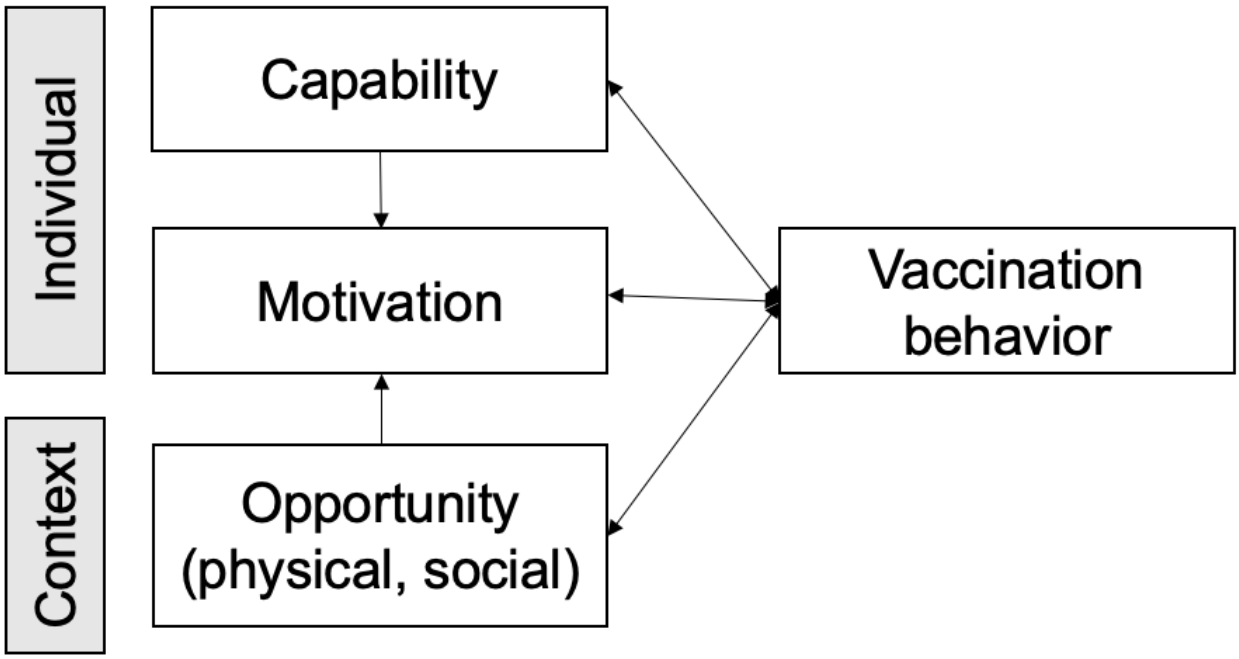
Previous studies have described potential impediments to COVID-19 vaccines, including questioning the need for vaccines and preferring to benefit from the immunity conferred by surviving COVID-19 [7]; safety issues regarding the rapid development and testing process of vaccines [7]; issues related to mandatory vaccination [7]; and conspiracy beliefs [8]. Some researchers conducted surveys based on theoretical models to explore facilitators and barriers to COVID-19 vaccination. Williams et al [9] conducted a survey to examine factors that influenced respondents' decisions to be vaccinated against COVID-19 and identified three facilitators (personal health, severity of COVID-19 disease, health consequences to others) and one barrier (concerns regarding vaccine safety). Lin et al [10] used the health belief model (HBM) to identify two facilitators (reduced likelihood of contracting COVID-10; others getting vaccinated) and one barrier (concerns about efficacy and side effects). Wong et al [11] also used the HBM to identify one facilitator of perceived benefits (the belief that the vaccination can reduce infection probability and alleviate concerns about COVID-19).

Compared with surveys, Twitter can gather timely information regarding behavioral intentions toward COVID-19 vaccines, especially to understand "anti-vaxxers" and those influenced by misinformation who are inclined not to get the vaccine. These types of users could be the most vulnerable population for COVID-19 vaccine outreach interventions. On the Twitter platform, there are more than 330 million users, and the median number of posts per person each month is 2 [12,13]. A recent survey of US Twitter users showed that they are younger and have a higher education level than the general population; however, their gender, race, and ethnicity distributions are similar to those of the general population [13]. The Twitter platform has been validated as a way to develop a public perception–tracking tool based on real-time content [14]. In addition, Twitter has been flooded with information about COVID-19, influenced by social isolation policies enacted during the epidemic [15]. The Twitter platform could be used to explore determinants of health-related behavior intentions [16]. Because the maximum length of each tweet is 280 characters, in addition to mentioning potential behavioral intentions, users can also briefly describe the reasons that led to their decision. In addition, using geotagged Twitter data, it is easier and faster to identify people's perceptions in different geographic locations. Therefore, we chose to use the Twitter platform to analyze behavioral intentions toward COVID-19 vaccines.

To better characterize the factors that influence behavioral intentions on the COVID-19 vaccines mentioned in the tweets, we conducted a theory-based content analysis based on the capability, opportunity, motivation–behavior (COM-B) model. The COM-B model was proposed by Michie et al [17] in 2011, and it contains three basic constructs: capability (physical and psychological), motivation (automatic and reflective), and opportunity (social and physical). The COM-B model is a comprehensive theoretical model based on causal mechanisms to identify individual and context factors that influence behavioral change. It was developed by merging 19 behavior change frameworks through a systematic literature review and discussions with behavior change experts. It has been successfully applied to many health-related behaviors, such as smoking cessation [18-20] and obesity reduction [21,22]. In contrast to other health behavior theories (eg, HBM, theory of planned behavior), the COM-B model was developed based on the behavior change wheel, which not only provides a theoretical analysis of behavior but, more importantly, provides results that can be used to assist with intervention design [17,23]. In addition, the World Health Organization (WHO) Regional Office for Europe has adapted it to vaccination behaviors to design its Tailoring Immunization Programmes (TIP) approach [24,25]. They merged subconstructs in capability and motivation, respectively. Because of vaccination behavior, the physical capability is interlinked with the psychological capability. Likewise, automatic motivation (ie, emotions, impulses) interacts with reflective motivation (ie, intentions, beliefs). Another advantage of the adapted COM-B model is that it focuses on vaccination behavior and provides refined details for each construct in the vaccine context. Therefore, we used an adapted COM-B model (see Figure 1) as the theoretical model of this study. The objectives of this study are to (1) determine if the adapted COM-B model can explain behavioral intentions toward COVID-19 vaccines using tweets; (2) examine theory-informed factors that may affect behavioral intentions toward COVID-19 vaccines; and (3) extract themes to provide information for public health researchers to develop theory-based and evidence-based promotion interventions.

XSL•FO
RenderX

Figure 1. The theoretical model based on the capability, opportunity, motivation—behavior (COM-B) model adapted to vaccination behavior, developed by the World Health Organization Regional Office for Europe [24,25].



Methods

Adapted COM-B Model

The adapted COM-B model has three theoretical constructs: (1) capability, (2) motivation, and (3) opportunity. The WHO Regional Office for Europe also provided examples for each construct (see Table 1). Capability refers to the individual's physical and psychological ability to perform the behavior, along with the knowledge and skills required to complete the activity [17]. In particular, psychological ability is the ability of an individual to have the necessary thought processes, such as being able to understand, reason, etc [17]. Motivation has a broad definition that includes goals and conscious decision-making as well as all other individual processes that motivate and lead to behavior, such as automatic processes (habitual processes, emotional responses) and reflective processes (analytical decision-making) [17]. Opportunity refers to the contextual factors that prompt the behavior to occur, including the physical opportunity and the environment and social opportunity [17]. The adapted COM-B model also provides a dynamic relationship between constructs. For example, both capability and opportunity can affect motivation. All three constructs of competence, motivation, and opportunity can generate behavior; on the other hand, behavior can influence these three constructs.

Table 1. The theoretical constructs in the adapted capability, opportunity, motivation–behavior (COM-B) model and associated examples by the World Health Organization Regional Office for Europe.

Theoretical construct	Examples
Capability	• Knowledge • Skills, trust in own skills • Resilience, stamina, will power, surplus energy • Physical fitness, ability
Motivation	• Attitudes, perceptions, risk assessment • Values, beliefs • Emotions, impulses, feelings • Confidence, trust
Opportunity (physical)	• Access, affordability, availability of vaccination • Convenience, appeal, appropriateness of vaccination • Structural efficiency • Availability of information
Opportunity (social)	• Social, cultural demands, support • Social, cultural cues, norms, values

XSL•FO
RenderX

Data Collection

We collected English tweets posted from November 1-22, 2020, using a combination of relevant keywords and hashtags: (#covid OR covid OR #covid19 OR covid19) AND (#vaccine OR vaccine OR #vacine OR vacine OR vaccinate OR immunization OR immune OR vax). After excluding retweets, we randomly selected 5000 tweets for manual coding and content analysis. The random numbers were generated through the NumPy package in Python. Then, we mapped the random numbers with the index of collected tweets.

Content Analysis

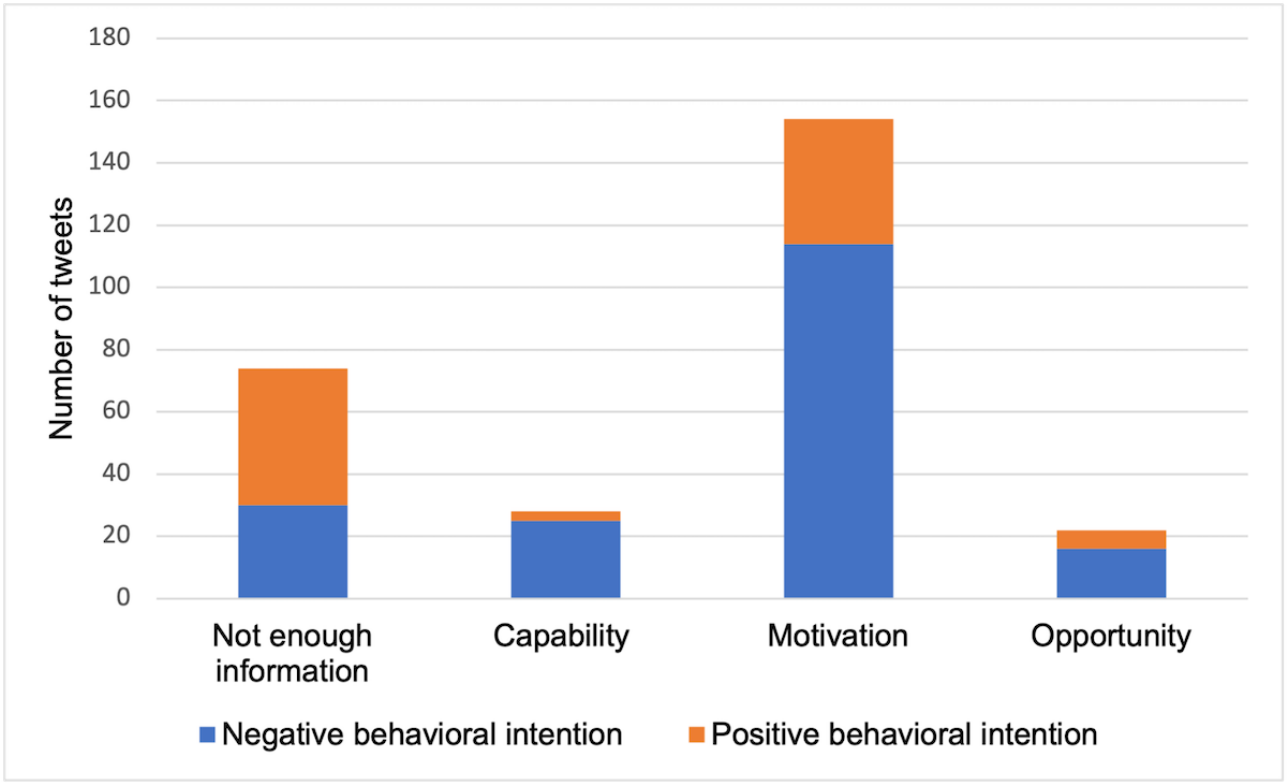
We performed a content analysis informed by the adapted COM-B model. The coding schema was developed iteratively. First, we developed the coding schema based on the definitions of constructs in the adapted COM-B model. Two reviewers (SL and JL) independently coded 1000 tweets in each round. After completing one round of coding, the two reviewers met with a third reviewer to discuss disagreements and update the coding schema until consensus was reached. We calculated the interrater reliability for the last round. If a tweet mentioned ≥2 constructs simultaneously, we coded it with multiple labels. We conducted chi-square tests to explore the relationship between theoretical constructs with the positive/negative behavioral intention. The statistical significance threshold was .05.

Results

Data Collection

We coded 5000 COVID-19 vaccine–related tweets, which were posted by 4796 unique users. We found 279 tweets that stated their behavioral intentions. The remaining tweets did not state any behavioral intentions toward COVID-19 vaccines. A total of 97 tweets were labeled with positive behavioral intentions, while 182 tweets contained negative behavioral intentions. Among them, 28 tweets were mapped with capability factors; 155 tweets were related to motivation; 23 tweets were related to opportunities; and 74 tweets did not contain any useful information about reasons for their behavioral intentions (see Figure 2). The κ value was 0.73. Of the tweets, 2 mentioned ≥2 constructs at the same time. Tweets that mentioned capability (χ^2_1=17.286, P<.001), motivation (χ^2_1=35.558, P<.001), and opportunity (χ^2_1=4.545, P=.03) were more likely to have negative behavioral intentions (Table 2).

Figure 2. Numbers of tweets containing different theoretical constructs.



XSL•FO
RenderX

Table 2. Themes and example tweets (n=204).

Theme	Value, n (%)	Example tweets
Capability (n=28)		
Knowledge	24 (11.8)	• "I am not getting a covid vaccine because there will be a microchip in there so the government can track my weekly shop in Lidls" • "No way I'm taking it. In one study out of women became steril [sic] from taking the vaccine. Is this the one that alters your DNA and can cause cancer? Thanks but I would rather get COVID." • "As I said it's not a 'from scratch' situation...covid is one version of a virus that a lot of work has already been done on. It's just a case of isolating the correct strain to vaccinate against. Look if you don't want to get it fine...all the more for me and my family."
Physical condition	4 (2)	• "I got my flu shot a few days ago. For now there is no way I will get a COVID vaccine." • "I have severe adverse reactions to vaccines and I meet of the exclusion criteria for the COVID vaccine trials."
Opportunity (n=22)		
Rights, regulation and legislation (eg, charging for vaccination, mandatory vaccination, inadequate regulation of vaccines, poor availability of information)	11 (5.4)	• "Tbh[a] I'm not against every vaccine in history. Just this specific COVID-19 vaccine. There are many reasons why, for example the nonsensical social regulations, that people like you probably want to be over so bad. But you seem to hot headed to absorb any legit information." • "If they charge us for this vaccine. I don't give. Not a single person should have to pay for a Covid Vaccine." • "Please stop volunteering me for this COVID-19 vaccine. I need more information. I've been close enough to death this year, no thank you. -a healthcare worker."
Social and cultural demands and support	9 (4.4)	• "Someone that works for the CDC[b] told me, do NOT get the covid vaccine.. Not sure but I'm not OK with her saying that at all." • "I'm not getting a covid vaccine. Do what yall want with that info to write me off. But its against my culture and everything I've been practicing."
Social consequences and reactions to vaccination	2 (1)	• "when anti vaccs realize everything is gonna start requiring the covid vaccine i can't wait to see. and for natural selection to do what it does best (unless you're immunocompromised and can't take certain vaccines, then they'll accept testing of course)"
Motivation (n=154)		
Attitudes and perceptions about the COVID-19 vaccine or the disease itself (eg, disease severity, vaccine effectiveness)	68 (33.3)	• "Yeah, with an average age of over you forgot to mention, a minute percentage of those that have had it and almost elderly and/or with underlying health conditions. For many of us, taking the rushed vaccine may more likely to cause us harm than covid itself" • "OK. so we each make our own risk/benefit decision. As the virus is what, % fatality, why would someone choose an unproven mRNA vaccine, with phase trials only since July? a biotech never before used programming RNA transcriptase to produce Covid in your cells? No thx."
Strong emotions about COVID-19 vaccine (eg, fear of side effects, unfairness to COVID-19 survivors)	5 (2.4)	• "We shouldn't have a COVID vaccine because it'll make those who lost loved ones to COVID angry."
Values and beliefs (eg, natural immune system, alternative medicine, value of vaccination)	37 (18)	• "My immune system is better than any vaccine. And a vaccine works in tandem with the way how your immune system beats virus No COVID Vax for me!" • "I had covid19 because I'm a long haul trucker and travel all over to different states..what I did to cure myself was take % alcohol and pour it on a rag and breathing in the fumes before I went to bed..I did that minutes a night..I got well in no time.. don't need no vaccine"

XSL•FO
RenderX

Theme	Value, n (%)	Example tweets
Confidence and trust (vaccine, health authorities, science, and government)	13 (6.3)	• "I do not trust the FDA[c], especially concerning the brand new Covid vaccine being rushed to market. Covid poses almost zero risk to my demographic. The pharmaceutical companies face zero risk if the vaccine damages their customers. For me the vaccine is riskier than the virus." • "I won't accept COVID vaccine offered by Biden administration..."

[a]Tbh: to be honest.

[b]CDC: US Centers for Disease Control and Prevention.

[c]FDA: US Food and Drug Administration.

Capability

The first theme regarding capability was that some users lacked knowledge about the COVID-19 vaccines and were influenced by misinformation. mRNA vaccines are a new technology, and some believe that these vaccines can alter DNA. As for the vaccine itself, some users believe that it contains a microchip that can be tracked by the government. Some people expressed that they did not know the side effects of COVID-19 vaccines or even their long-term effects. Based on this concern, misinformation was generated that the vaccine could cause sterility, cancer, etc. Other reasons for not getting the vaccine included the belief that the vaccine could make people sick, the belief that there is no need for people who have been infected with COVID-19 to get the vaccine, and that the vaccine does not prevent infection but only alleviates symptoms. All these beliefs reflect a lack of understanding of the COVID-19 vaccines among users. Having more knowledge about vaccines could help people develop positive behavioral intentions; for example, some people mentioned that they could understand why the development process was fast because researchers only needed to isolate the correct strain rather than create one. Others mentioned understanding that vaccines do not contain live viruses, so they would be willing to get the vaccine.

Some users emphasized that their physical condition was not suitable for COVID-19 vaccination and that they were unsure how their body would react after vaccination (eg, recently had an influenza vaccine shot, had a suppressed immune system, had a severe adverse reaction to vaccination in the past, had a stroke).

Opportunity

In the physical opportunities category, some users said they would not accept the vaccine if they had to pay for it. Many users said vaccination should be a free choice and that they would refuse to receive the vaccine if it became mandatory. Some users wanted more regulation of the COVID-19 vaccine, and some were concerned about the availability of information.

Among the social opportunities, we found some factors influencing the COVID-19 vaccination related to social and cultural demands and support, such as going against religion, defying culture, health workers not recommending vaccination, and family members actively discouraging vaccination. Others intended to be vaccinated because of the fear of the social consequences and reactions to vaccination, such as fear of affecting their work and requiring proof of vaccination for many activities in the future.

Motivation

Most of the reasons for behavioral intentions were categorized into the motivation construct. Many users expressed attitudes and opinions about the COVID-19 vaccine or the disease itself, such as not considering the disease to be severe or life-threatening and not considering the vaccine to be effective (because of low efficiency and mutation of the virus). Others assessed the risks and considered the rushed vaccine to be more harmful than COVID-19. Some of those with positive behavioral intentions stated that they chose vaccination because they did not want to be infected with COVID-19.

Some users expressed strong emotions and feelings about COVID-19 vaccines, such as fear of the vaccine and concern about its side effects because it is new. Notably, others felt that vaccination was selfish or unfair, as they would not be exposing themselves to the same risk as others who survived COVID-19; thus, they were reluctant to be vaccinated.

Other themes were values and beliefs. Some users expressed a belief that the body's natural immune system is better than any vaccine, or they believed more in alternative medicine. On the other hand, others who wanted to be vaccinated emphasized the positive values of vaccination, such as saving more lives, reopening the economy, and returning to normal life.

Confidence and trust were dominant themes. Users expressed distrust in many areas: the quality of the vaccines (hastily manufactured, untested), health authorities, science, companies, and government. For Twitter users in the United States, we found that unlike with other vaccines, part of the reason people do not trust the government is because of their past handling of the COVID-19 pandemic. In addition, the COVID-19 vaccine rollout came during a time when a new president was being elected in the United States, and some people lacked confidence in the newly elected president's party.

Discussion

Principal Findings

In this study, we conducted a theory-based content analysis using a dataset of 5000 tweets posted from November 1-22, 2020. We identified 279 tweets that contained behavioral intentions regarding COVID-19 vaccines and mapped them to constructs in the adapted COM-B model. We generated nine themes that influence Twitter users' intentions to receive COVID-19 vaccines. The constructs in the COM-B model could be applied systematically to characterize factors that influence behavioral intentions toward COVID-19 vaccines. In addition,

XSL•FO
RenderX

we found that among tweets that simply stated behavioral intentions without including any reason, the number of positive intention tweets was higher than that of negative intention tweets. The results also implied that more than half of the tweets expressing the decision-making process were negative intention tweets. This finding aligns with our expectation of understanding the factors that contribute to vaccine hesitancy to better develop tailored vaccine promotion programs.

The novelty of COVID-19 vaccines and the current social context have created further difficulties in vaccine rollout. Identified barriers of influenza vaccination intention and behavior include lack of confidence (eg, negative attitudes, mistrust), inconvenience (eg, cost, access), calculation (eg, risk assessment), and complacency (eg, underestimating disease severity) [26]. Our results revealed the presence of several other factors that influence behavior toward COVID-19 vaccination. First, misinformation or conspiracy theories about COVID-19 and COVID-19 vaccines are much more prevalent on social media than those about other diseases or their related vaccines [27-29]. Some users were influenced by this misinformation, and this led to refusal of the vaccine. Second, Twitter users expressed concerns about mandatory vaccination. At the Emergency Use Authorization stage, mandatory vaccination is legally and ethically questionable [30]. However, with full Biologics License Application approval, policy makers may mandate vaccination for all populations. Given the existence of users who have strongly indicated that they would not accept mandatory vaccination against COVID-19, policy makers must be cautious in determining vaccination policies for the public. Some studies have suggested that vaccine mandates do not improve vaccine acceptance rates, and a proposed alternative approach is to apply informed risk communication and give people the freedom to choose without compromising personal autonomy [30]. Third, users with positive behavioral intentions emphasized that the positive value of vaccination to society, such as restoring economic and normal life from before the epidemic, motivated them to be vaccinated. This facilitator is uncommon in other vaccination behaviors. This facilitator could be matched with a strategy of converting personal decisions into public acts [31]. Fourth, we observed that some users were reluctant to be vaccinated because they felt that the COVID-19 vaccine was unfair to those who survived. Based on this concern, it may be helpful to select COVID-19 survivors as opinion leaders to promote COVID-19 vaccines, increasing public awareness of the severity and risk of the disease. Fifth, in addition to the mistrust of vaccines and science that also exists for other vaccines, for the COVID-19 vaccine, users in the United States expressed more mistrust of the government for two specific reasons: (1) the previous administration's inappropriate behavior in handling COVID-19 and (2) the lack of confidence in the newly inaugurated president's political party. These findings were also aligned with previous studies that proposed a role of politics in COVID-19 vaccine hesitancy [8]. Sixth, it is worth noting that even though previous studies have identified that past vaccination behavior can be used to predict future vaccination behavior [32], we found that past experiences with other vaccines may not affect COVID-19 vaccination. For example, one user mentioned:

> *I'm not against every vaccine in history. Just this specific COVID-19 vaccine. There are many reasons why, for example the nonsensical social regulations.*

The above differences contribute to the fact that the rollout of COVID-19 vaccines could be more complicated than that of other vaccines. Researchers need to develop interventions specific to the COVID-19 vaccine to improve acceptance rates. This also provides an opportunity for future studies to comprehensively analyze why behavioral intentions toward COVID-19 vaccines are different from those toward other vaccines.

In several studies, Twitter content analyses have been conducted of health care behaviors other than COVID-19 vaccination. For example, Chew and Eysenbach [14] collected 2009 H1N1-related tweets and identified the resource content posted most often, followed by personal experience, personal opinion, jokes, marketing, and spam. Li et al [33] extracted COVID-19 stigma–related tweets and found that group labeling, responsibility, and peril tweets disseminated the stigma. Furthermore, several studies have validated the usability of tweets through theory-based content analysis to promote breast cancer promotion programs [16,34]. Our study is the first study to analyze the behavior intention of COVID-19 vaccines through a theory-based content analysis using social media content.

Limitations

This study has several limitations. First, we only analyzed the behavioral intentions of users on Twitter. Previous studies have shown that health care providers are the primary advocates for vaccination and largely influence vaccination acceptance rates [35-38]. In this platform, we could not distinguish users' occupations. However, the themes reported in this study could help researchers to develop evidence-based interventions for the general public. To examine the vaccine behavior of health care providers, we will conduct a questionnaire-based survey, and the results from that study could aid the development of clinical guidelines for health workers. This approach of considering the general public separately from health care providers is also recommended by the TIP approach developed by the WHO Regional Office for Europe [24]. Second, there are differences between behavioral intentions and actual vaccine behaviors. However, behavioral intentions have been shown to directly influence actual behaviors [39-41]. Third, Twitter users are considered to be younger and have a higher level of education than the general public [13]. Based on this concern, further qualitative research could be conducted on the older population, those with lower education levels, or those with limited access to the internet.

In future research, a literature review could be conducted to summarize current implementation strategies for COVID-19 vaccine promotion and map them to the themes identified in this study to determine gaps in recent research. The inner mechanism of the adapted COM-B model—the behavior change wheel—could inform evidence-based and theoretical implementation strategies to improve the effectiveness of COVID-19 vaccine promotion programs.

XSL•FO
RenderX

Conclusion

The study demonstrates the capability of applying the COM-B model to characterize behavioral intentions toward COVID-19 vaccines on the Twitter platform. We successfully generated nine themes of factors that affect behavioral intentions. Positive behavioral intentions were affected by the positive values of vaccination (eg, reduced risk of infection, socioeconomic recovery, return to normal life). In contrast, negative behavioral intentions were associated with attitudes and perceptions about COVID-19 vaccines or the disease itself (eg, underestimation of disease severity, low vaccine effectiveness), values and beliefs (eg, greater belief in the natural immune system), confidence and trust (eg, distrust of government or vaccines), and lack of knowledge. The generated themes could be used to create theory-based and evidence-based implementation strategies to promote COVID-19 vaccines.

Authors' Contributions

JL and SL conceived the study. SL and JL performed the analysis, interpreted the results and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-01-31]
2. Liu J, Liu S. The management of coronavirus disease 2019 (COVID-19). J Med Virol 2020 Sep;92(9):1484-1490 [FREE Full text] [doi: 10.1002/jmv.25965] [Medline: 32369222]
3. Liu J, Liu S, Wei H, Yang X. Epidemiology, clinical characteristics of the first cases of COVID-19. Eur J Clin Invest 2020 Oct 11;50(10):e13364. [doi: 10.1111/eci.13364] [Medline: 32725884]
4. Pogue K, Jensen J, Stancil C, Ferguson DG, Hughes SJ, Mello EJ, et al. Influences on Attitudes Regarding Potential COVID-19 Vaccination in the United States. Vaccines (Basel) 2020 Oct 03;8(4):582 [FREE Full text] [doi: 10.3390/vaccines8040582] [Medline: 33022917]
5. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
6. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477 [FREE Full text] [doi: 10.3201/eid2607.200282] [Medline: 32255761]
7. Schaffer DeRoo S, Pudalov NJ, Fu LY. Planning for a COVID-19 vaccination program. JAMA 2020 Jun 23;323(24):2458-2459. [doi: 10.1001/jama.2020.8711] [Medline: 32421155]
8. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]
9. Williams L, Gallant AJ, Rasmussen S, Brown Nicholls LA, Cogan N, Deakin K, et al. Towards intervention development to increase the uptake of COVID-19 vaccination among those at high risk: outlining evidence-based and theoretically informed future intervention content. Br J Health Psychol 2020 Nov 05;25(4):1039-1054. [doi: 10.1111/bjhp.12468] [Medline: 32889759]
10. Lin Y, Hu Z, Zhao Q, Alias H, Danaee M, Wong LP. Understanding COVID-19 vaccine demand and hesitancy: A nationwide online survey in China. PLoS Negl Trop Dis 2020 Dec 17;14(12):e0008961 [FREE Full text] [doi: 10.1371/journal.pntd.0008961] [Medline: 33332359]
11. Wong LP, Alias H, Wong P, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother 2020 Sep 01;16(9):2204-2214. [doi: 10.1080/21645515.2020.1790279] [Medline: 32730103]
12. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. Statista. 2021 Jan 27. URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2021-04-03]
13. Wojcik S, Hughes A. Sizing Up Twitter Users. Pew Research Center. 2019 Apr 24. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/2019/04/twitter_opinions_4_18_final_clean.pdf [accessed 2021-05-06]
14. 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]
15. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. ArXiv. Preprint posted online on July 5, 2020 [FREE Full text]
16. Diddi P, Lundy LK. Organizational Twitter use: content analysis of tweets during Breast Cancer Awareness Month. J Health Commun 2017 Mar 19;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
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]

XSL•FO
RenderX

18. Gould GS, Bar-Zeev Y, Bovill M, Atkins L, Gruppetta M, Clarke MJ, et al. Designing an implementation intervention with the Behaviour Change Wheel for health provider smoking cessation care for Australian Indigenous pregnant women. Implement Sci 2017 Sep 15;12(1):114 [FREE Full text] [doi: 10.1186/s13012-017-0645-1] [Medline: 28915815]
19. Raupach T, Falk J, Vangeli E, Schiekirka S, Rustler C, Grassi MC, et al. Structured smoking cessation training for health professionals on cardiology wards: a prospective study. Eur J Prev Cardiol 2014 Jul 24;21(7):915-922. [doi: 10.1177/2047487312462803] [Medline: 23008136]
20. Fulton E, Brown K, Kwah K, Wild S. StopApp: using the behaviour change wheel to develop an app to increase uptake and attendance at NHS Stop Smoking Services. Healthcare (Basel) 2016 Jun 08;4(2):31 [FREE Full text] [doi: 10.3390/healthcare4020031] [Medline: 27417619]
21. Croker H, Russell SJ, Gireesh A, Bonham A, Hawkes C, Bedford H, et al. Obesity prevention in the early years: a mapping study of national policies in England from a behavioural science perspective. PLoS One 2020 Sep 30;15(9):e0239402 [FREE Full text] [doi: 10.1371/journal.pone.0239402] [Medline: 32997681]
22. Flannery C, McHugh S, Anaba AE, Clifford E, O'Riordan M, Kenny LC, et al. Enablers and barriers to physical activity in overweight and obese pregnant women: an analysis informed by the theoretical domains framework and COM-B model. BMC Pregnancy Childbirth 2018 May 21;18(1):178 [FREE Full text] [doi: 10.1186/s12884-018-1816-z] [Medline: 29783933]
23. Gould GS, Bar-Zeev Y, Bovill M, Atkins L, Gruppetta M, Clarke MJ, et al. Designing an implementation intervention with the Behaviour Change Wheel for health provider smoking cessation care for Australian Indigenous pregnant women. Implement Sci 2017 Sep 15;12(1):114 [FREE Full text] [doi: 10.1186/s13012-017-0645-1] [Medline: 28915815]
24. Tailoring Immunization Programmes: TIP. World Health Organization. 2019. URL: https://www.euro.who.int/en/health-topics/disease-prevention/vaccines-and-immunization/publications/2019/tip-tailoring-immunization-programmes-2019 [accessed 2021-05-06]
25. Dubé E, Leask J, Wolff B, Hickler B, Balaban V, Hosein E, et al. The WHO Tailoring Immunization Programmes (TIP) approach: review of implementation to date. Vaccine 2018 Mar 07;36(11):1509-1515 [FREE Full text] [doi: 10.1016/j.vaccine.2017.12.012] [Medline: 29287678]
26. Schmid P, Rauber D, Betsch C, Lidolt G, Denker M. Barriers of influenza vaccination intention and behavior - A systematic review of influenza vaccine hesitancy, 2005 - 2016. PLoS One 2017 Jan 26;12(1):e0170550 [FREE Full text] [doi: 10.1371/journal.pone.0170550] [Medline: 28125629]
27. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
28. Evanega S, Lynas M, Adams J, Smolenyak K. Coronavirus misinformation: quantifying sources and themes in the COVID-19 'infodemic'. JMIR Preprints. Preprint posted online on October 19, 2020. [doi: 10.2196/preprints.25143]
29. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: detecting COVID-19 misinformation in Twitter. IEEE Access 2020;8:155961-155970. [doi: 10.1109/access.2020.3019600]
30. Gostin LO, Salmon DA, Larson HJ. Mandating COVID-19 vaccines. JAMA 2021 Feb 09;325(6):532-533. [doi: 10.1001/jama.2020.26553] [Medline: 33372955]
31. Volpp KG, Loewenstein G, Buttenheim AM. Behaviorally informed strategies for a national COVID-19 vaccine promotion program. JAMA 2021 Jan 12;325(2):125-126. [doi: 10.1001/jama.2020.24036] [Medline: 33315079]
32. Lin CJ, Nowalk MP, Toback SL, Rousculp MD, Raymund M, Ambrose CS, et al. Importance of vaccination habit and vaccine choice on influenza vaccination among healthy working adults. Vaccine 2010 Nov 10;28(48):7706-7712. [doi: 10.1016/j.vaccine.2010.07.009] [Medline: 20638452]
33. Li Y, Twersky S, Ignace K, Zhao M, Purandare R, Bennett-Jones B, et al. Constructing and communicating COVID-19 stigma on Twitter: a content analysis of tweets during the early stage of the COVID-19 outbreak. Int J Environ Res Public Health 2020 Sep 19;17(18) [FREE Full text] [doi: 10.3390/ijerph17186847] [Medline: 32961702]
34. Lyles CR, López A, Pasick R, Sarkar U. "5 mins of uncomfyness is better than dealing with cancer 4 a lifetime": an exploratory qualitative analysis of cervical and breast cancer screening dialogue on Twitter. J Cancer Educ 2013 Mar;28(1):127-133. [doi: 10.1007/s13187-012-0432-2] [Medline: 23132231]
35. Leask J, Kinnersley P, Jackson C, Cheater F, Bedford H, Rowles G. Communicating with parents about vaccination: a framework for health professionals. BMC Pediatr 2012 Sep 21;12:154 [FREE Full text] [doi: 10.1186/1471-2431-12-154] [Medline: 22998654]
36. Cooper LZ, Larson HJ, Katz SL. Protecting public trust in immunization. Pediatrics 2008 Jul 01;122(1):149-153. [doi: 10.1542/peds.2008-0987] [Medline: 18595998]
37. Schmitt H, Booy R, Aston R, Van Damme P, Schumacher RF, Campins M, et al. How to optimise the coverage rate of infant and adult immunisations in Europe. BMC Med 2007 May 29;5(1):11 [FREE Full text] [doi: 10.1186/1741-7015-5-11] [Medline: 17535430]
38. Habersaat KB, Jackson C. Understanding vaccine acceptance and demand-and ways to increase them. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Jan 04;63(1):32-39 [FREE Full text] [doi: 10.1007/s00103-019-03063-0] [Medline: 31802154]

XSL•FO
RenderX

39. Williams MD, Rana NP, Dwivedi YK. The unified theory of acceptance and use of technology (UTAUT): a literature review. J Enterp Inf Manag 2015 Apr 13;28(3):443-488. [doi: 10.1108/JEIM-09-2014-0088]
40. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control. SSSP Springer Series in Social Psychology. Berlin, Germany: Springer; 1985.
41. Hale J, Householder B, Greene K. The theory of reasoned action. In: The Persuasion Handbook: Developments in Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2002:259-286.

Abbreviations

COM-B: capability, opportunity, motivation–behavior
HBM: health belief model
TIP: Tailoring Immunization Programmes
WHO: World Health Organization

Edited by C Basch; submitted 22.02.21; peer-reviewed by C Huang, R Cochran, EM Schomakers; comments to author 03.04.21; revised version received 10.04.21; accepted 29.04.21; published 12.05.21.

Please cite as:
Liu S, Liu J
Understanding Behavioral Intentions Toward COVID-19 Vaccines: Theory-Based Content Analysis of Tweets
J Med Internet Res 2021;23(5):e28118
URL: https://www.jmir.org/2021/5/e28118
doi:10.2196/28118
PMID:33939625

©Siru Liu, Jialin Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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

Influence of COVID-19 Lockdowns on the Usage of a Vision Assistance App Among Global Users With Visual Impairment: Big Data Analytics Study

Gang Luo[1], PhD; Shrinivas Pundlik[1], PhD

Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Gang Luo, PhD
Schepens Eye Research Institute
Massachusetts Eye and Ear
Harvard Medical School
20 Staniford Street
Boston, MA, 02114
United States
Phone: 1 617 912 2529
Email: gang_luo@meei.harvard.edu

Abstract

Background: Millions of individuals with visual impairment use vision assistance apps to help with their daily activities. The most widely used vision assistance apps are magnifier apps. It is still largely unknown what the apps are used for. Lack of insight into the visual needs of individuals with visual impairment is a hurdle for the development of more effective assistive technologies.

Objective: This study aimed to investigate how needs for visual aids may vary with social activities, by observing the changes in the usage of a smartphone magnifier app when many users take breaks from work.

Methods: The number of launches of the SuperVision Magnifier app was determined retrospectively from 2018 to 2020 from among active users worldwide. The fluctuation in app usage was examined by comparing weekday vs weekend periods, Christmas and new year vs nonholiday seasons, and COVID-19 lockdowns vs the easing of restriction during the pandemic.

Results: On average, the app was used 262,466 times by 38,237 users each month in 2020 worldwide. There were two major trough points on the timeline of weekly app usage, one aligned with the COVID-19 lockdowns in April 2020 and another aligned with the Christmas and new year week in 2018 and 2019. The app launches declined by 6947 (11% decline; $P<.001$) during the lockdown and by 5212 (9% decline; $P=.001$) during the holiday weeks. There was no significant decline during March to May 2019. App usage compensated for seasonal changes was 8.6% less during weekends than during weekdays ($P<.001$).

Conclusions: The need for vision assistance technology was slightly lower during breaks and lockdowns, probably because the activities at home were different and less visually demanding. Nevertheless, for the entire user population, the needs for visual aids are still substantial.

(J Med Internet Res 2021;23(5):e26283) doi:10.2196/26283

KEYWORDS

assistance; assistive technology; COVID-19; development; eye; low vision; needs; smartphone apps; usage; vision assistance; vision; visual impairment

Introduction

People with impaired visual acuity typically use magnifying devices for a wide range of daily reading tasks. As smartphones and tablets have become popular in recent years, there has been an increase in the use of vision assistance mobile apps, which utilize embedded cameras and turn mobile devices into handheld electronic magnifiers. As the mobile devices are already at hand, their convenience, together with the benefits of electronic magnification, such as an adjustable zoom level, have provided significant advantages over traditional optical magnifiers. The mobile apps are increasingly used by young [1] and older individuals with visual impairment [2]. A recent survey reported that >50% of people with visual impairment use smartphone magnification apps [3]. Our recent study with more than 16,000 magnifier app users reported that they mostly use the app for

XSL•FO
RenderX

short spot-reading tasks [4]. This finding is consistent with a view accepted among those with low vision: magnifier use is not limited to reading only long-passage text. Some studies have focused on spot reading in activities of daily living, such as finding phone numbers and reading instructions on medicine bottles [5-8].

However, what constitutes the spectrum of activities for which people are using mobile magnifier apps is still an open question. Understanding the purposes of app use can be significant for low vision research, because technology development and evaluation ideally should respond to the actual, real-life needs of users. Otherwise, a visual aid, although reportedly useful, might not be as useful in everyday life.

The important knowledge about the needs of individuals with visual impairment for vision assistance include, albeit not exclusively, the purpose of visual tasks (eg, mobility and leisure reading), visual targets (eg, text, object, and pet), environment (eg, location and lighting conditions), effectiveness (or gap), and the availability of vision assistance means (eg, magnifiers and sighted companions). It should be noted that their needs may evolve dynamically. Starke et al [9] recently classified the visual needs of 32 people with low vision on the basis of images captured by body-mounted cameras over a 1-week period. They reported that individuals with visual impairment frequently need help with using mobile devices, which were not widely available 10 years ago.

This study reports an important finding that can contribute to our understanding of the visual needs of smartphone magnifier app users. Specifically, this study aimed to investigate whether vision assistance app users are those who are mostly confined at home owing to vision impairment. Our goal is to find evidence indicating that some of app usages are involved in social activities outside of home.

Instead of a direct approach to interview users, this study harnessed a global event, the COVID 19 pandemic, as an unexpected intervention to deduce the underlying purpose of magnifier app uses. The approaches to directly observe users' behaviors can result in less ambiguous data, but it is usually very challenging to enroll a large sample of participants owing to the potential introduction of a sampling bias. Our chosen method studies the entire app user population as a black box. While the study has a limitation with regard to the lack of an experimental control, obtaining big data from tens of thousands of users has the potential to reveal larger population-level trends and may yield valuable findings.

Specifically, if a significantly large number of magnifier app users are not confined at home under normal circumstances and go to work or school, app usage is expected to decrease during the COVID-19 lockdown. On the contrary, if the magnifier apps are mostly used for household activities at home, app usage is not expected to decrease but rather increase during the lockdown, as people spend more time at home. As this study is related to social activity, we also investigated app usage during weekends and holidays, with the same expectation of observing the same effect as that during the lockdown.

Methods

SuperVision Magnifier App and Users

Data presented here were collected for 38 weeks, from November 2019 to July 2020, using the freely available SuperVision Magnifier app, which was developed by our group. The iOS version of the SuperVision Magnifier (version 1.7.16 used in this study) was released worldwide in October 2016. The app provides vision assistance features, which are commonly available in other smartphone magnification apps and dedicated handheld video magnifiers: up to 16× digital zoom, the ability to capture a snapshot, color inversion (and binarizing images), and a toggling flashlight. The app can be considered a representative of many other similar magnifier apps available in the App store.

Data were collected from active users worldwide, which included 92% of returning users, who repeatedly used the app at different frequencies, and 8% of new users, who installed the app for the first time. Because the log data of 1 app launch are always uploaded to the server when the app is opened the next time, those new users must have used the app at least twice. A new user becomes a returning user if he/she launches the app at least 1 day after the first time. People who installed the app but did not use it during the period of data collection were not included in this study.

Data Collection Methods

A data collection module is embedded in the app to collect runtime usage data. Each time the app became the foreground app, a launch event was recorded. The data logging module was developed using the analytics software development kit of Umeng (Beijing, China). Data were received in a consolidated form from the whole user group, and the users' privacy was not compromised as no participant-identifiable information was collected. The study was approved by the institutional review board of the Massachusetts Eye and Ear Infirmary and was exempt from the regulations of studies on human subjects because the aggregated data prevented the identification of participants.

To further obtain direct evidence regarding the geolocations where the app is used, we released a new version (1.8.1) in September 2020, in which the distance between 2 consecutive app launches is calculated and uploaded to the Umeng server. The specific geolocation data are not saved because of the requirement of privacy protection. These types of data would indicate whether the app is used only at the same location.

For user activity analysis, consolidated weekly app usage reports were downloaded from the website of the analytics service provider, Umeng. For the timeline of major events related to the COVID-19 pandemic, information was extracted from several reliable websites [10-15].

Statistical Analysis

Statistical analysis was performed using the time series data for global weekly app launches from November 2019 to July 2020. Since the time series observations are not independent, autoregressive integrated moving average (ARIMA) modeling

XSL•FO
RenderX

was used for analysis. The effect of COVID-19 lockdowns on magnifier app usage was analyzed using an ARIMA model, which included a predictor that indicated the presence of an intervention; that is, the occurrence of COVID-19 lockdowns. A similar but separate analysis was conducted to determine the effect of holidays on app usage.

The following segments of the weekly app launch data time series were identified: (1) “Holidays”: a 2-week duration related to Christmas and new year’s day; (2) “Prelockdown”: the weeks before March 15, 2020, excluding holidays, which were assumed to represent normal app usage in the absence of interventions; (3) “Lockdown”: 8 weeks starting from the week of March 15 to May 10, 2020; and (4) “Postlockdown”: the weeks after May 10, 2020. Given that the pandemic affected different countries and regions worldwide at slightly different times, the rationale for choosing the period between March 15 and May 10, 2020, as the lockdown period was that these 8 weeks coincided with the spread of the pandemic across a large part of the world, including North America, Europe, and Asia.

The effects of lockdown and holiday interventions on app usage were analyzed separately. When analyzing the effect of holidays, only the prelockdown and holidays segments of the time series were considered. When analyzing the effect of lockdowns, only the prelockdown and lockdown segments were considered.

Statistical analysis was performed using the ARIMA function from the standard library in R (The R Foundation), along with associated functions for evaluating and diagnosing the fitted model. Coefficient (β) values, SE, and *P* values are reported when describing the effect of the interventions. An effect with a *P* value of <.05 was considered significant.

Because the active users of the app grew over time, when investigating app usage during weekends, the number of daily app launches was normalized by the weekly average to exclude the effect of the growing trend. This measure could also compensate for the seasonal changes because of holidays and COVID-19 lockdowns in this study.

Results

Between November 2019 and July 2020, the app was used 262,466 (SD 10,277) times by 38,237 (SD 983) users each month and 60,505 (SD 2,877) times by 14,585 (SD 474) users each week on average, worldwide. On average, each user launched the app 4.1 (SD 0.1) times per week. The solid curve in Figure 1 shows the fluctuation in weekly app launches from November 2019 to July 2020. There are 2 major troughs in the curve, which appear to be aligned with the new year holiday and COVID-19 lockdowns. The number of app uses decreased from the preholiday weekly peak of 63,807 to 56,510 during the holiday weeks, and it decreased from the prelockdown peak of 66,182 to the lowest point at 54,702 during the lockdowns. As some countries started to ease lockdown restrictions, the weekly app launches rose to 58,730 in mid-May 2020, and the recovery appears to have continued through July 2020. To exclude the possibility that the decline during the lockdown period was due to a seasonal change and not lockdowns, the weekly app launches from November 2018 to July 2019 are also plotted as a dotted curve in Figure 1. As shown in Figure 1, there was a decline in app launches around the new year holiday of 2019, but there was no major decline in April 2019. This observation suggests that the decline during April 2020 was not seasonal but rather possibly due to the COVID-19 lockdowns.

An ARIMA (0,1,1) model with interventions was fitted to the weekly app launch data. The effect of lockdowns on app use was found to be significant (β=–6947, SE=1039; *P*<.001). The negative value of the coefficient for the lockdown variable indicates that the number of app launches decreased by an average of 6947 per week (approximately 11%) during the lockdown compared to the prelockdown phase. Similarly, the intervention resulting from holidays had a significant effect (β=–5212, SE=1307; *P*=.001). Again, the negative value of the coefficient indicates that the weekly average app launches decreased by 5212 during the holiday weeks.

From November 2018 to July 2019, there was a significant reduction in app launches around the holiday weeks of Christmas and new year’s day (β=–2168, SE=720; *P*=.003). However, there was no significant change in the number of app launches per week during mid-March to early May 2019 (β=–643, SE=603; *P*=.29).

Since the pandemic’s course was not the same across all countries, the fluctuation in app usage was also visualized by country or region to confirm that the patterns observed in the global data could still be observed in each country or region. App users were located in more than 120 countries, but most of the countries had only a small number of users. We selected four countries or regions with most users for further analysis: Russia, the United States, Japan, and Europe (Figure 2). There were more than 1500 active users weekly in each of these selected regions. No other country had an average of more than 588 weekly users.

XSL•FO
RenderX

Figure 1. Weekly app launches (number of times of app usage) from November 2019 to July 2020, compared to the previous year's data for the same duration (dotted line). For the curve representing November 2019 to July 2020, there are 2 troughs around new year's day and COVID-19 lockdowns in April 2020. A corresponding trough near new year's day 2019 is observed in the previous year's curve, but there is no corresponding decline in April 2019. G_a: approximately one-third of the world's population is living with COVID-19 restrictions, G_b: some countries, including Spain, Iran, Italy, Denmark, Israel, Germany, New Zealand, and Thailand began to ease lockdown restrictions.

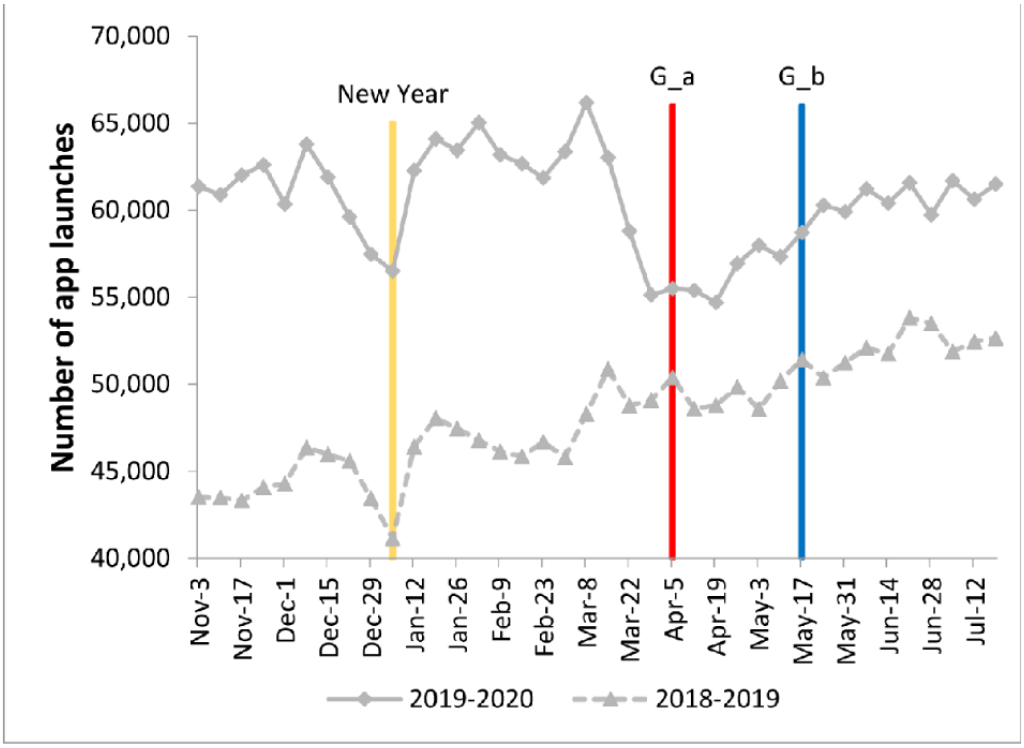
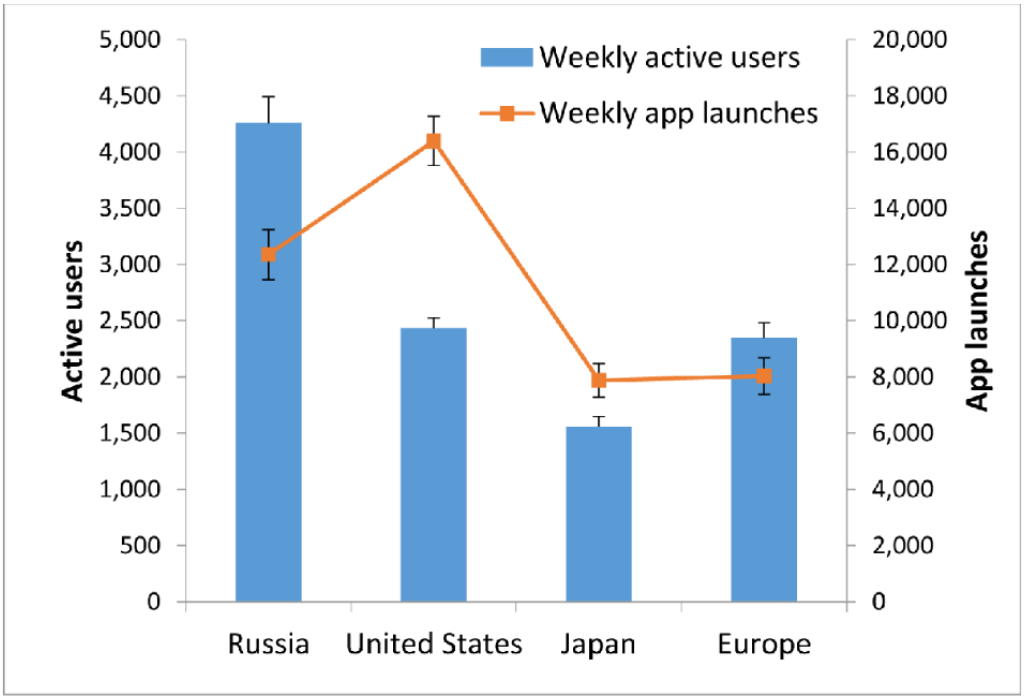


Figure 2. Four regions were included for separate analysis: Russia, the United States, Japan, and Europe. Their average numbers of weekly active users were 4259, 2436, 1558, and 2346, respectively. The corresponding average numbers of weekly app launches (number of times the app was used) were 12,353, 16,393, 7880 and 8031, respectively. Error bars indicate SD values.



Figures 3-6 show the fluctuation in weekly app launches during a period of 38 weeks from November 2019 to July 2020 in the aforementioned 4 regions. The timeline of some major events are superimposed on the fluctuation curves. Three types of events are color-coded: yellow=holiday, red=lockdown, and blue=easing of lockdown restrictions. It can be seen that Christmas or new year are aligned with one of the major troughs in the 4 regions. For the United States, there seemed to be an additional holiday trough around Thanksgiving.

XSL•FO
RenderX

Figure 3. App launch fluctuations in Russia, normalized by the average number of weekly app uses. Yellow: holiday, red: lockdown, and blue: postlockdown. In Russia, Eastern Orthodox Christmas was on January 7. R_a: Ministry of Culture announced the closure of all cultural institutions under its jurisdiction; Minister of Education Sergey Kravtsov announced the closure of all Russian schools from March 23, 2020. R_b: Moscow issued a stay-at-home order for all residents. R_c: President Vladimir Putin announced an end to Russia's 6-week nationwide lockdown despite a record increase in COVID-19 cases.

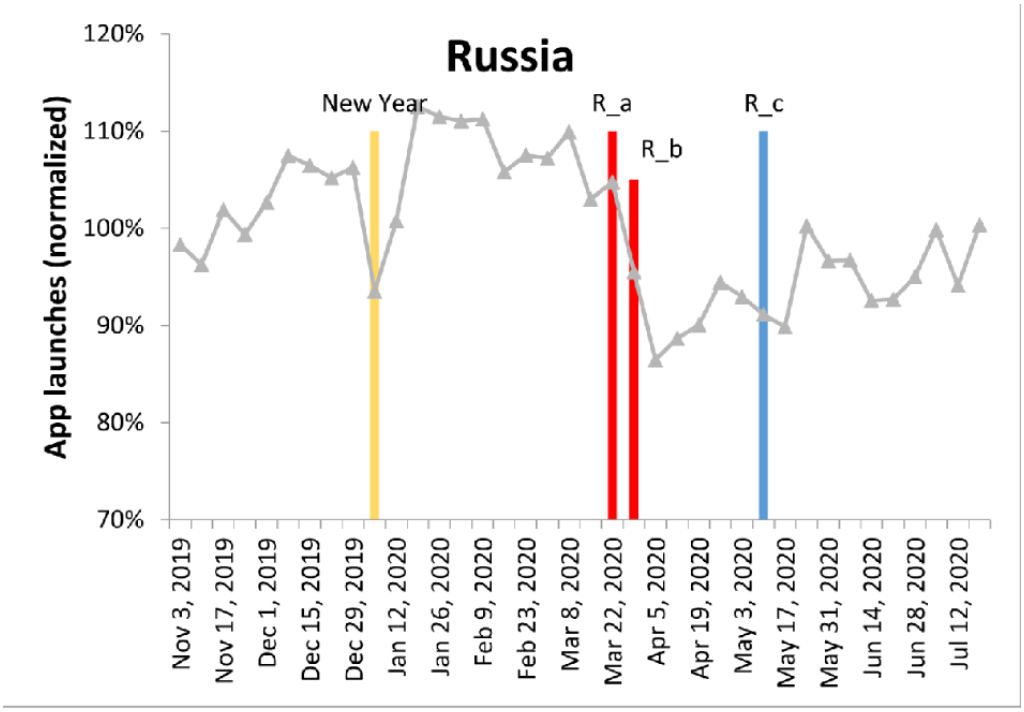
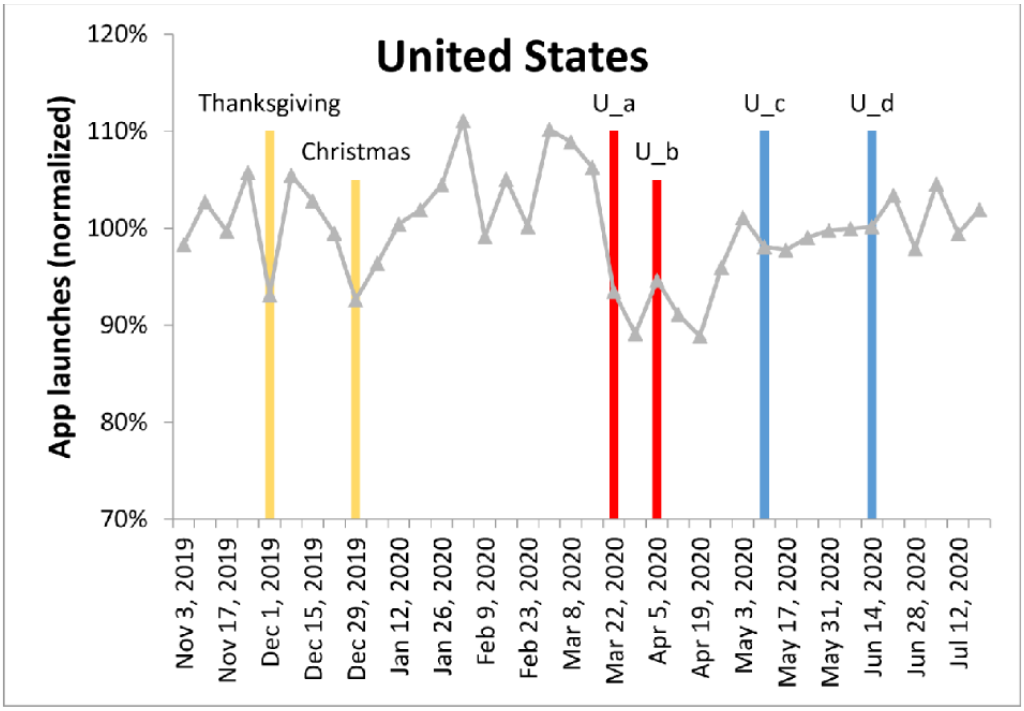


Figure 4. App launches in the United States, color-coded as in Figure 3. U_a: New York City public schools closed, and California issued a stay-at-home order for all 40 million residents. U_b: approximately 91% of people in the United States were instructed to stay at home. U_c: Florida and California reopened. U_d: New York City began phase I of reopening.



XSL•FO
RenderX

Figure 5. App launches in Japan, color-coded as in Figure 3. J_a: Japan's prime minister proclaimed a 1-month state of emergency from April 8 to May 6, 2020, for Tokyo and the prefectures of Kanagawa, Saitama, Chiba, Osaka, Hyogo, and Fukuoka. J_b: Japan's prime minister lifted the state of emergency imposed in 39 of 47 prefectures, announcing that the nation's rate of infection has decreased to one-seventh of its peak.

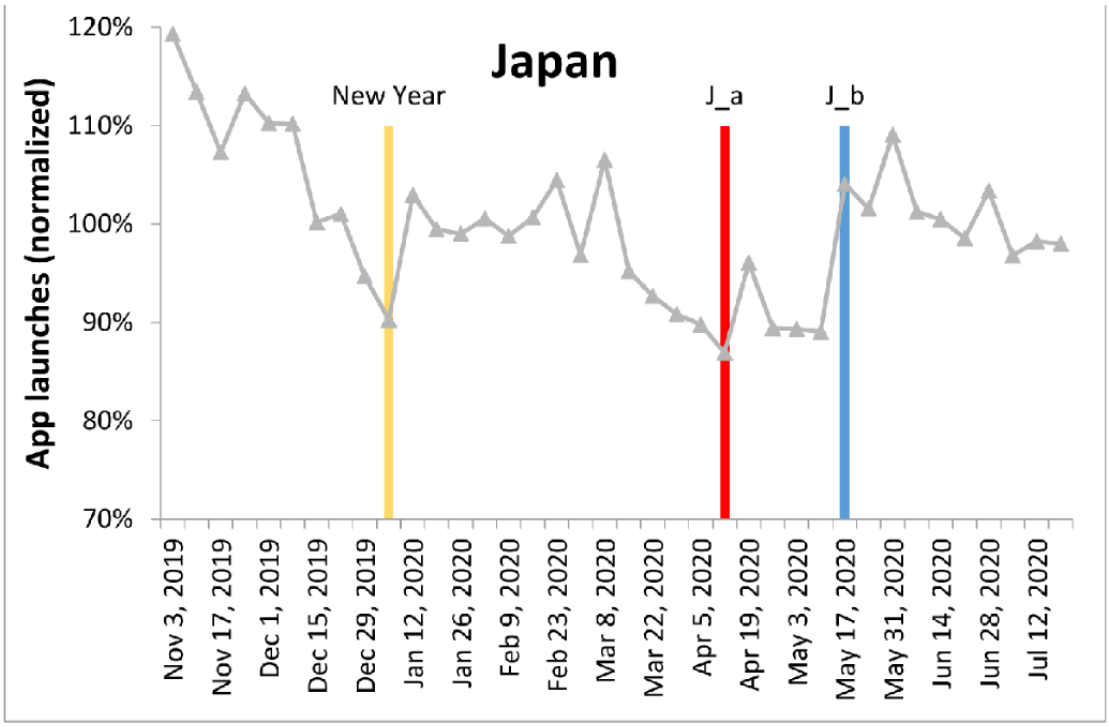
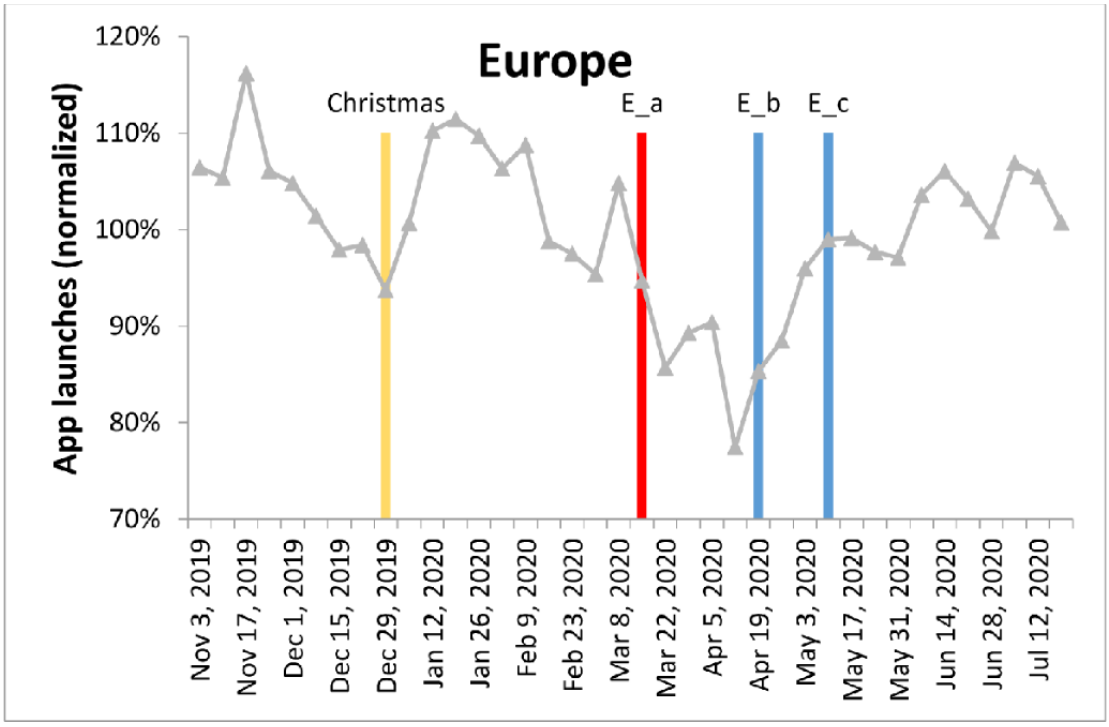


Figure 6. App launches in Europe, color-coded as in Figure 3. E_a: All shops and venues closed nationwide in Italy. E_b: Economy reopened in Germany. E_c: Lockdown restrictions started to be eased in several countries such as Spain, Italy, Denmark, and Germany.



Consistently, the lockdown events in these 4 regions were either during the major trough in the global data or during the decreasing phase shortly before that trough. Similarly, the regional lockdown easing events occurred during the increasing phase of global app use.

Figure 7 shows the normalized daily app usage by days of the week between November 2019 and July 2020. Overall, app usage on weekends was 8.6% lesser than that on weekdays (93.9% vs 102.5%, respectively; $P<.001$). As shown in Figure 7, app usage was the lowest on Sunday. App usage on Saturday and Monday, which are similar to that on Sunday, were also lower than the that during other weekdays. Because app users were at different time zones worldwide, the local time for most users differed from the time of the Umeng analytic server (GMT+8 hours). Therefore, the effect of the weekend spills

XSL•FO
RenderX

over 1 day before and 1 day after. If Saturday and Monday are excluded, usage on Sunday was 12% lesser than the average usage from Tuesday to Friday.

Figure 8 shows the distribution of the distance between 2 consecutive app launches. The data were collected across a period from September 15, 2020, to February 16, 2021, including total of 414,618 samples. Far from being restricted to 1 place, a small number of users travelled more than 500 km. Most consecutive app uses were recorded at the same locations, but this does not necessarily indicate that 1 pair of (consecutive) app launches occurred at the same location as that of another pair of (consecutive) app launches. A sizeable peak was observed between 5 km and 50 km (7.2%). For reference, the average commute distance in the United States is 24 km each way [16].

Figure 7. Daily app launches normalized by the weekly average for the period between November 2019 and July 2020. Bars indicate mean app usage by the days of the week. Error bars indicate SD values.

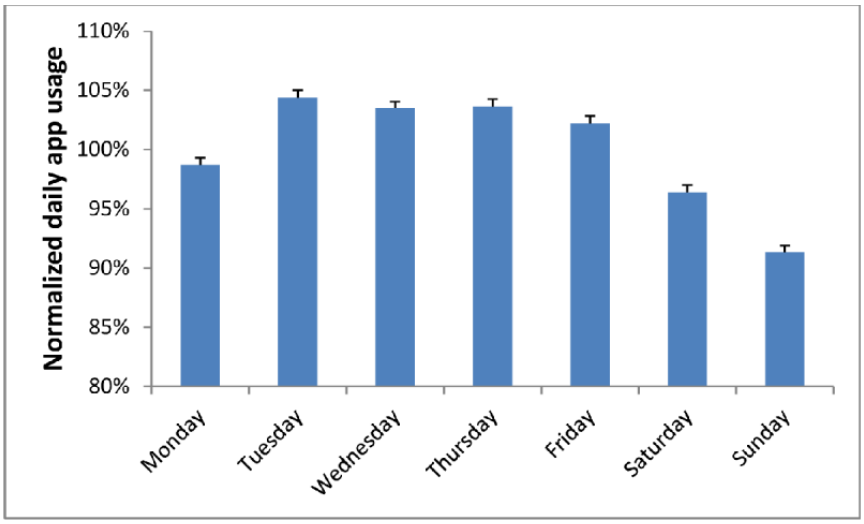
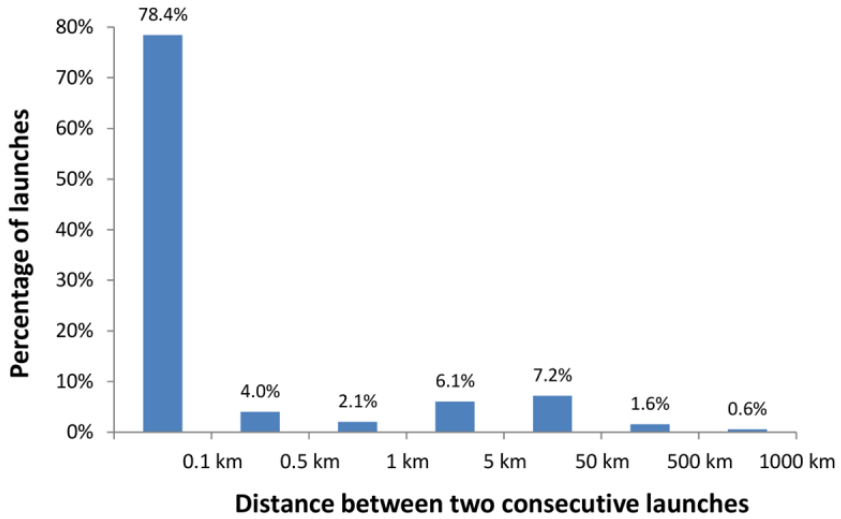


Figure 8. Percentage of app launches vs the distance to the location of the previous launch. Most consecutive app launches occurred at the same place. Excluding the bin for the same location, there is a sizeable peak between 5 km and 50 km, which indicates that 7.2% of the consecutive app launches were separated by a distance of 5-50 km.



Discussion

Principal Findings

This study investigated the impact of changes in social activities on the app launch pattern of tens of thousands of SuperVision Magnifier app users and observed a declining trend in app usage corresponding to weekends, major holidays, and the initiation of lockdowns in different regions compared to the prelockdown usage frequency. Given the black box of the mixed user population, the global COVID-19 lockdown effectively provided a window through which insights can be gained by investigating the effect of a behavioral intervention, which is not planned or designed. In this study, this opportunity was utilized to help deduce the situations where the vision assistance app was used.

Our previous study reported that the mobile magnifier app is mostly used for short spot-reading, while a small number of users may use the app for long passage–reading for ≥30 minutes each time [4]. In terms of app launch, a small portion of users used the app many times on a daily and weekly basis, and many users used it only occasionally. It should be noted that the number and duration of app uses are not necessarily a measure

XSL•FO
RenderX

of helpfulness. Those metrics are typically used to gauge the customer adherence to video game and social media apps. For vision assistance apps, those metrics reflect the needs of users (especially returning users). For instance, Google Maps is very useful for traveling to unfamiliar places, but many people do not use it every day. In addition to launch and duration measures, our latest study on visual targets viewed by users through the app reported that more than half of the targets are nontextual targets including indoor and outdoor contexts, food, animals, and plants [17]. Furthermore, our study shows that the locations, where vision assistance is needed, are not restricted to home. Together, the wide range of frequency, the variety of target categories, and the effects of social activity–related locations suggest that the magnifier app can probably meet many different visual needs in the daily lives of people with visual impairment. Our ongoing studies are attempting to generate an overview of the visual needs of these individuals from different perspectives.

It remains unknown what factors influence the needs or in what context those visual tasks are performed with help from the app. Without fully understanding the actual visual needs in the real world, limited studies on low vision and technologic developments have been relying on subjective opinions and limited surveys. Thus far, previous studies have evaluated the spot reading performance of vision assistance interventions based on visual tasks that are typically carried out at home [5,7], including reading information on medicine bottles, product labels, and utility bills, all of which are textual targets. However, visual tasks outside of home (eg, at work, school, or during traveling) have been rarely considered. If we assume that the magnifier app is mostly used at home, app use would have increased during lockdowns; however, we observed the opposite trend from the app usage data.

The data on distance between 2 consecutive app launches directly indicate that the magnifier app is also frequently used outside home. Interestingly, the travel distance seemed to be in line with the average commute distance in the United States. While our study population was large, detailed information on geolocations could not be collected. However, considering a consistent reduction in app usage globally during weekends, holidays, and COVID-19 lockdowns, it is a reasonable speculation that the reduction could be due to reduced work- or school-related activities. If this association is accurate, it would suggest that normally the magnifier app is used not only at home for activities of daily living but also at workplaces, schools, or while traveling. Regarding the proportion of app usage away from home, this study could not collect direct data. The 11% decline during lockdowns and 8.6% of decline on weekends might be a rough indicator of the frequency of such usage. According to anecdotal reports we received, the app was used by some people to read road signs or to watch presentations in conferences. However, these are only a few examples of app use outside of home. More precise estimation will be conducted in future studies.

A counter argument might be made that the reduction in app usage during COVID-19 lockdowns may not necessarily be due to a reduction in work- or school-related activities but instead, it could be simply because the users had less chances to use the app when staying at home during lockdowns. This study did not identify the environment in which the app was used, but the hypothesis cannot explain the reductions during holidays and weekends. The time people spend outside their home for traveling, socializing, outing, and shopping during these breaks should be more than that during COVID-19 lockdowns. More activities outside home during weekends in 2019 and Christmas and new year holiday weeks in 2018 and 2019 did not appear to increase app usage. On the other hand, if the decline during holidays and weekends were to be explained by less time spent at home, assuming that the magnifier app is primarily used for activities of daily living at home, then the explanation based on the time staying at home would not be able to account for the decline during lockdowns. Taken together, a more plausible explanation could be that decreased app usage during weekends, holidays, and lockdowns are due to reduced work- or school-related activities carried out away from home. As many people worked from home and many students practice remote learning during the lockdowns, it is possible some users still use the vision assistance app at home to help with their work. This study could not distinguish this portion from that of nonwork- or nonschool-related app usages. If this were true, it would imply that the app is normally used even outside home when there is no lockdown.

Another possible reason for the reduction in app usage may be that when users with visual impairment were at home, their family members and friends could help with some of the activities of daily living, or the visual tasks at home was less demanding overall than those at workplaces; hence, they did not need to use the app in some cases. Nevertheless, for the whole user population, the need for vision assistance devices on a daily basis is still substantial, as the reduction in app usage during breaks was only approximately 10%.

To confirm the aforementioned speculations, future studies are required to collect direct evidence indicating the detailed nature of activities for which the app was used, and to conduct some necessary surveys (eg, on demographics and vision status) to help interpret the behavioral data. Further big data analysis studies will help reveal the wide spectrum of visual demands for vision assistance among people with visual impairment. This knowledge will be important for developing personalized, more effective solutions in future low vision rehabilitation research. The development of vision assistance technologies can target a particular group of users (eg, elderly individuals) or the entire cohort. Whatever the targeted group is, ideally, the composition of visual tasks should roughly match the spectrum of visual needs in the targeted cohort's actual visual needs. Otherwise, studies may miss important needs while over-representing others. Without accurate feedback from appropriate evaluation studies, the utility of technologies may be compromised.

Conclusions

Considering our large user sample, this longitudinal study observed a slight reduction in the usage of the SuperVision Magnifier app during weekends, holidays, and COVID-19 lockdowns. We surmise that the possible reasons could be that the magnifier app is normally used in some work- and school-related activities in addition to daily tasks performed at

XSL•FO
RenderX

home, and the visual tasks at home may be relatively less demanding than those performed outside of home.

Acknowledgments

This study is supported in part by a National Institutes of Health award (award# P30EY003790). We thank Anurag Shubaham for upgrading the app.

Conflicts of Interest

GL is involved in the development of the vision app, which is completely free to the public. He has no financial competing interest related to the app.

References

1. Chun R, Bhakhri R, Coalter J, Jay WM. Smartphone Usage in Patients with Optic Atrophy. Neuro-Ophthalmology 2012 Sep 25;36(5):193-195. [doi: 10.3109/01658107.2012.710923]
2. Bhakhri R, Chun R, Coalter J, Jay W. A Survey of Smartphone Usage in Low Vision Patients. Invest Ophthalmol Vis Sci 2012;53(14):4421.
3. Crossland MD, Silva RS, Macedo AF. Smartphone, tablet computer and e-reader use by people with vision impairment. Ophthalmic Physiol Opt 2014 Sep;34(5):552-557. [doi: 10.1111/opo.12136] [Medline: 25070703]
4. Luo G. How 16,000 people used a smartphone magnifier app in their daily lives. Clin Exp Optom 2020 Nov;103(6):847-852. [doi: 10.1111/cxo.12996] [Medline: 31773801]
5. Wittich W, Jarry J, Morrice E, Johnson A. Effectiveness of the Apple iPad as a Spot-reading Magnifier. Optom Vis Sci 2018 Sep;95(9):704-710 [FREE Full text] [doi: 10.1097/OPX.0000000000001269] [Medline: 30153239]
6. Pundlik S, Singh A, Baghel G, Baliutaviciute V, Luo G. A Mobile Application for Keyword Search in Real-World Scenes. IEEE J Transl Eng Health Med 2019;7:2900210 [FREE Full text] [doi: 10.1109/JTEHM.2019.2935451] [Medline: 31667028]
7. Owsley C, McGwin G, Sloane ME, Stalvey BT, Wells J. Timed instrumental activities of daily living tasks: relationship to visual function in older adults. Optom Vis Sci 2001 May;78(5):350-359. [doi: 10.1097/00006324-200105000-00019] [Medline: 11384013]
8. West SK, Rubin GS, Broman AT, Muñoz B, Bandeen-Roche K, Turano K. How does visual impairment affect performance on tasks of everyday life? The SEE Project. Salisbury Eye Evaluation. Arch Ophthalmol 2002 Jun;120(6):774-780. [doi: 10.1001/archopht.120.6.774] [Medline: 12049583]
9. Taylor DJ, Hobby AE, Binns AM, Crabb DP. How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. BMJ Open 2016 Dec 02;6(12):e011504 [FREE Full text] [doi: 10.1136/bmjopen-2016-011504] [Medline: 27913556]
10. Neilson S, Woodward A. A comprehensive timeline of the coronavirus pandemic at 1 year, from China's first case to the present. Insider. 2020 Dec 24. URL: https://www.businessinsider.com/coronavirus-pandemic-timeline-history-major-events-2020-3 [accessed 2021-05-07]
11. Kantis C, Kiernan S, Bardi J. UPDATED: Timeline of the Coronavirus. Think Global Health. 2021 Mar 26. URL: https://www.thinkglobalhealth.org/article/updated-timeline-coronavirus [accessed 2021-05-07]
12. COVID-19 pandemic in Japan. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Japan [accessed 2021-05-10]
13. Xue F. Japan's Response to the Coronavirus: A Timeline of Major Events. University of Virginia: College and Graduate School of Arts and Sciences. 2020 May 18. URL: https://eastasiacenter.as.virginia.edu/news/japans-response-coronavirus-timeline-major-events [accessed 2021-05-07]
14. COVID-19 pandemic in Russia. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Russia [accessed 2021-05-10]
15. Harding L. Global report: Russia to end national lockdown as French shops reopen. The Guardian. 2020 May 11. URL: https://www.theguardian.com/world/2020/may/11/russia-to-end-national-covid-19-lockdown-as-cases-hit-record-high [accessed 2021-05-07]
16. From Home to Work, the Average Commute is 26.4 Minutes. OmniStats. 2003. URL: https://www.nrc.gov/docs/ML1006/ML100621425.pdf [accessed 2021-05-07]
17. Luo G. What Visual Targets Are Viewed by Users With a Handheld Mobile Magnifier App. Trans Vis Sci Tech 2021 Mar 01;10(3):16. [doi: 10.1167/tvst.10.3.16]

Abbreviations

ARIMA: autoregressive integrated moving average

XSL•FO
RenderX

Edited by C Basch; submitted 05.12.20; peer-reviewed by A Colenbrander, L Kremer; comments to author 09.02.21; revised version received 25.02.21; accepted 15.04.21; published 12.05.21.

Please cite as:
Luo G, Pundlik S
Influence of COVID-19 Lockdowns on the Usage of a Vision Assistance App Among Global Users With Visual Impairment: Big Data Analytics Study
J Med Internet Res 2021;23(5):e26283
URL: https://www.jmir.org/2021/5/e26283
doi:10.2196/26283
PMID:33945492

©Gang Luo, Shrinivas Pundlik. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2021. 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 Efficacy of Computerized Cognitive Behavioral Therapy for Depressive and Anxiety Symptoms in Patients With COVID-19: Randomized Controlled Trial

Zhifen Liu[1], MD; Dan Qiao[1], MM; Yifan Xu[1], MM; Wentao Zhao[1], MAP; Yang Yang[1], MSN; Dan Wen[1], MM; Xinrong Li[1], MD; Xiaoping Nie[2], MBBS; Yongkang Dong[3], MBBS; Shiyou Tang[4], MBBS; Yi Jiang[5], MD; Ying Wang[6], MSN; Juan Zhao[1], MSN; Yong Xu[1], MD

[1]Department of Psychiatry, The First Hospital of Shanxi Medical University, Taiyuan, China

[2]Department of Medical Service, Chongqing Public Health Medical Center, Chongqing, China

[3]The Fourth People's Hospital of Taiyuan, Taiyuan, China

[4]Department of Psychiatry, Chongqing University Three Gorges Hospital, Chongqing, China

[5]Department of Respiratory and Critical Disease Medicine, The First Hospital of Shanxi Medical University, Taiyuan, China

[6]Department of Geriatrics, The First Hospital of Shanxi Medical University, Taiyuan, China

Corresponding Author:
Yong Xu, MD
Department of Psychiatry
The First Hospital of Shanxi Medical University
85 Jiefang South Street, Yingze District
Taiyuan, 030001
China
Phone: 86 351 4639641
Email: xuyongsmu@vip.163.com

Abstract

Background: The prevalence of depressive and anxiety symptoms in patients with COVID-19 is higher than usual. Previous studies have shown that there are drug-to-drug interactions between antiretroviral drugs and antidepressants. Therefore, an effective and safe treatment method was needed. Cognitive behavioral therapy (CBT) is the first-line psychological therapy in clinical treatment. Computerized CBT (cCBT) was proven to be an effective alternative to CBT and does not require face-to-face therapy between a therapist and the patient, which suited the COVID-19 pandemic response.

Objective: This study aims to evaluate the efficacy of the cCBT program we developed in improving depressive and anxiety symptoms among patients with COVID-19.

Methods: We customized a cCBT program focused on improving depressive and anxiety symptoms among patients with COVID-19, and then, we assessed its effectiveness. Screening was based on symptoms of depression or anxiety for patients who scored ≥7 on the Hamilton Depression Rating Scale ($HAMD_{17}$) or the Hamilton Anxiety Scale (HAMA). A total of 252 patients with COVID-19 at five sites were randomized into two groups: cCBT + treatment as usual (TAU; n=126) and TAU without cCBT (n=126). The cCBT + TAU group received the cCBT intervention program for 1 week. The primary efficacy measures were the $HAMD_{17}$ and HAMA scores. The secondary outcome measures were the Self-Rating Depression Scale (SDS), Self-Rating Anxiety Scale (SAS), and Athens Insomnia Scale (AIS). Assessments were carried out pre- and postintervention. The patients' symptoms of anxiety and depression in one of the centers were assessed again within 1 month after the postintervention assessment.

Results: The cCBT + TAU group displayed a significantly decreased score on the $HAMD_{17}$, HAMA, SDS, SAS, and AIS after the intervention compared to the TAU group (all P<.001). A mixed-effects repeated measures model revealed significant improvement in symptoms of depression ($HAMD_{17}$ and SDS scores, both P<.001), anxiety (HAMA and SAS scores, both P<.001), and insomnia (AIS score, P=.002) during the postintervention and follow-up periods in the cCBT + TAU group. Additionally, the improvement of insomnia among females (P=.14) and those with middle school education (P=.48) in the cCBT + TAU group showed no significant differences when compared to the TAU group.

XSL•FO
RenderX

Conclusions: The findings of this study suggest that the cCBT program we developed was an effective nonpharmacological treatment for symptoms of anxiety, depression, and insomnia among patients with COVID-19. Further research is warranted to investigate the long-term effects of cCBT for symptoms of anxiety, depression, and insomnia in patients with COVID-19.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000030084; http://www.chictr.org.cn/showprojen.aspx?proj=49952

(J Med Internet Res 2021;23(5):e26883) doi:10.2196/26883

KEYWORDS

mental health; depression; anxiety; COVID-19; treatment; cCBT; computerized cognitive behavioral therapy

Introduction

In the past decades, the effects of physical and psychological distress have increased with each successive public health emergency, such as with the severe acute respiratory syndrome (SARS) in 2003 [1], the Middle East Respiratory Syndrome (MERS) [2] in 2012, the Ebola virus disease in 2014 [3], and COVID-19 in 2020 [4]. COVID-19, caused by a novel coronavirus (SARS-CoV-2) [5], continues to spread worldwide. On January 30, 2020, the World Health Organization Emergency Committee declared a global health emergency based on growing case notification rates at Chinese and international locations [6]. At present, there have been nearly 30 million confirmed cases and nearly 1 million deaths worldwide [7].

The COVID-19 epidemic is similar to SARS and MERS [8] but has spread more quickly and efficiently [9]. In the acute stage of SARS and MERS, the prevalence of psychological distress among confirmed patients was 63.0%; common symptoms included insomnia (41.9%), anxiety (35.7%), and depression (32.6%) [10]. A cross-sectional study on psychological distress in patients with COVID-19 showed that the prevalence of symptoms of depression and anxiety was 35.9% and 38.5%, respectively [11].

The common clinical symptoms of patients with COVID-19 include cough, fever, fatigue, hypoxia, and occasionally gastrointestinal infection [12]. Moreover, the patients experienced various stressors during isolation, including but not limited to drug side effects, a fear of severe disease consequences, a fear of infecting others, and an inability to get correct information in a timely manner [13]—all of which may lead to several psychological distresses including anxiety, depression, and insomnia [8,14]. A lack of timely intervention in psychological distresses may affect the quality of life for patients in the future. As a survey of SARS survivors showed, 42.5% were diagnosed with at least one psychiatric disorder 4 years following the outbreak, with the most common diagnoses being posttraumatic stress disorder (PTSD; 54.5%) and major depressive disorder (39.0%). Further, 40.3% of survivors experienced chronic fatigue [15].

Early, brief, and trauma-focused mental health services may have the potential to reduce or delay the development of adverse outcomes [16]. An effective method may be cognitive behavioral therapy (CBT), which is widely used to treat mild to moderate depression and anxiety [17]. However, traditional CBT requires a face-to-face implementation, which is inconvenient for patients with COVID-19 who are quarantined in hospitals. Computerized CBT (cCBT), as an alternative mode of delivery, can make up for this deficiency. There is substantial evidence that cCBT is an effective alternative to CBT for the treatment of mild to moderate depression and anxiety [18-20]. Compared to CBT, cCBT offers a range of benefits. cCBT can be accessed with a computer, smartphone, or tablet without the need for face-to-face therapy, and this is especially key during a pandemic to reduce the risk of spreading infection [21]. cCBT does not require appointments, allowing instead for time flexibility and cost-effectiveness [22]. The privacy of patients is better protected, and the stigma of face-to-face therapy is mitigated [23].

Thus, to solve the psychological problems that may occur during the epidemic, we independently developed a cCBT program for patients with COVID-19. The program is a brief self-guided, online psychological intervention that consists of cognitive training, cognitive consolidation, and behavioral therapy via video and animation content, all of which is easy to understand and operate. It may comprehensively regulate mental distress that includes anxiety, depression, panic, and fear among patients with COVID-19; it may also help patients build confidence as they confront the disease.

Moreover, evidence from epidemiological studies revealed the relationship between gender, age, education level, and severity of mental problems [24]. Females and relatively high educational level were found to be predictors of mental distress of adults 1 month in the COVID-19 epidemic [25]. Additionally, several studies also suggest differences between men and women in response mechanisms to psychological distress. Men tend to *externalize* their distress by directing action outward, while women tend to *internalize* their distress by directing action inward [26,27]. Such gender difference may influence the treatment response [28]. Other studies have shown that the patients' age and educational level at the beginning of treatment may affect the final outcomes [29].

Based on all these factors, we conducted a multicenter, randomized, placebo-controlled clinical trial to evaluate the efficacy of the cCBT program. The program was developed with the aims of improving the symptoms of depression and anxiety among patients with COVID-19. We hypothesized greater symptom improvement with respect to mitigating anxiety and depression in the intervention group compared to the treatment as usual (TAU) group at both the postintervention and follow-up periods. We also hypothesized that online self-help may not be suitable for all patients, and the overall responses in mental problems may vary by gender, age, and educational level.

XSL•FO
RenderX

Methods

Trial Design

This was a prospective, multicenter, two-arm (allocation ratio was 1:1), parallel group; further, it was a nonblinded randomized controlled trial that was registered at the Chinese Clinical Trial Registry with the study number ChiCTR2000030084. Participants were recruited from five hospitals in China from March 2020 to June 2020. Eligible participants were screened for symptoms of anxiety or depression. Patients with symptoms of anxiety, depression, or both were included because these two symptoms often occur simultaneously. Patients were then randomly allocated to a 1-week course of the cCBT program we developed in addition to TAU (cCBT + TAU group), or they were randomly allocated to the control condition, the TAU group, when they completed the baseline assessment. The outcomes of both groups were measured twice—at the baseline and post intervention. To investigate the relative continuous effects of cCBT on symptoms of depression and anxiety, the outcomes of both groups in one of the sites were measured again 1 month after the postintervention.

Participants and Recruitment

The participants were recruited from the following five hospitals: Affiliated Tongji Hospital of Huazhong University of Science and Technology, Wuhan Institute for Tuberculosis Control, Central Hospital of Chongqing Three Gorges, Chongqing Public Health Medical Center, and the Fourth People's Hospital of Taiyuan. The inclusion criteria those who were aged between 18-75 years (including threshold), regardless of gender; were under isolation observation while diagnosed with a mild or common type of COVID-19 and could cooperate to complete a corresponding psychological intervention; had mild to moderate depressive or anxiety symptoms as defined by the 17-item Hamilton Depression Rating Scale ($HAMD_{17}$) score≥7 or the Hamilton Anxiety Scale (HAMA) score≥7; had sufficient compliance to complete the experiment according to the protocol; and informed consent was provided by patients and (if necessary) guardians. The exclusion criteria were patients who were clearly diagnosed with a psychiatric disorder, including depression, bipolar disorder, etc, in the 6 months prior to their diagnosis of COVID-19; patients with psychotic symptoms; patients with $HAMD_{17}$ score≥24 or HAMA score≥21; patients with a high risk of suicide, defined as having a history of attempts by suicide in the 6 months prior to the study or who scored more than 3 on item three (suicide item) of the $HAMD_{17}$ scale; patients with organic mental disorders; patients with substance abuse or dependence; patients undergoing treatment currently (pharmacological or psychological) for mental health problems; or patients presenting other conditions that the researchers believed were not suitable for this clinical trial.

The researcher explained to the eligible patients they had been introduced to the study because they were experiencing mental symptoms related to the burden of COVID-19 and that the intelligent therapy via the cCBT program we developed may help to reduce these distresses. Patients were also told that the aim of the study was to find out whether receiving this intelligent therapy might help them to relieve these mental symptoms linked to COVID-19. Participants were informed that they had a 50% chance of receiving the cCBT intervention and that the study would be offered as an additional source of support for free, not as a replacement or restriction to their current physical care. Besides, they would be provided with free medical consultation for a period of 3 years after discharge.

Randomization

Resman, a public management platform for clinical trial management of China's clinical trial registration agency, was used to carry out random grouping; the allocated sequence was kept by the administrator. Patients with mild to moderate anxiety or depression were randomized to the cCBT + TAU group or to the TAU group at a 1:1 ratio. Randomization was balanced using a block size of four and was stratified by institution.

Intervention

The cCBT Program

The program is a remote intervention model based on cCBT that we developed. The system can systematically intervene in patients' cognition, emotions, and behavior through an offline mobile terminal. It is easy to understand and operate, so patients can complete the work with the help of nonpsychological professionals or with self-guidance. Participants in the cCBT + TAU group were informed of how to access the program and use it during the study after registration.

Compared to other forms of digital CBT for anxiety and depression, the most unique feature of the cCBT program we developed is its targeting of patients with COVID-19. Components of the intervention were presented to patients with COVID-19 with anxiety and depressive symptoms through computer-based, visually attractive, and interactive examples, exercises, and videos. Given that the focus of CBT is on patients' irrational cognition. The first module of cCBT, called the cognitive therapy module, aims to minimize or even eliminate patients' negative thoughts about COVID-19. This part is presented in videos to participants. It mainly includes rational cognition of COVID-19, stress status during isolation, and psychosomatic mechanisms under stress in addition to sleeping management and education about status post isolation.

To further deepen the correct cognition of patients, we set up a cognitive consolidation module. Unlike the first module, this part is presented in the form of a miniature game, which undoubtedly increases the flexibility and interest of the intervention. The participants are required to answer questions based on videos related to the cognitive therapy module.

For the last part, the behavioral therapy module is targeted at teaching methods of regulating negative emotions. This module is also presented in the video to guide participants in relaxation training. Information about the following three relaxation training methods was provided:

1. Relaxation mental imagery training: The patients are taught to subjectively follow their own ideas and imagine some pleasant situations such as prairies, streams, trees, and greenery of native mountains. During the whole training process, patients are completely in a relaxed state with

XSL•FO
RenderX

gentle and even breathes. With the vivid image gradually becoming clear in their mind, the patients will feel a warm current running through the whole body. It is during these moments that patients can now enter a deep state of relaxation benefiting both body and mind.

2. Mindfulness meditation: The goal of this training is to improve the ability of self-regulation. Using the training of concentration and relaxation, it requires the patients to consciously maintain attention on the current internal or external experience without making any judgments about it.
3. Counting meditation: Patients with COVID-19 are generally unable to relax due to the fear of the highly contagious disease. Based on this, the training guides the patients to focus on the breath while meditating. The patient's physical and mental state will gradually calm down, and the ability to focus will continue to improve during this period of training.

This experimental treatment was delivered through more than 10 minutes of self-directed individual therapy per day for 1 week at each trial center. The program is installed on an iPad and is only available to research therapists. After the therapist shows the patients how to use the system, the patients can begin their journey of *self-help intervention*.

Treatment as Usual

TAU consisted of periodic psychological assessments, general psychological support, and consultations discussing overall well-being and disease activity. Patients whose assessment results suggested a certain risk to themselves or others at the time of each assessment were to be treated by a professional psychiatrist and were withdrawn from this study.

Measures

Data Collection Procedure

Demographic and clinical characteristics including age, gender, and education levels (years) were collected at baseline. $HAMD_{17}$ and HAMA scores were determined by trained researchers at the baseline, postintervention, and 1-month follow-up (hosted by only one of the centers). Secondary outcome measures were also recorded at the same point. These measures included the Self-Rating Depression Scale (SDS), the Self-Rating Anxiety Scale (SAS), and the Athens Insomnia Scale (AIS).

Outcome

Before the study, a questionnaire was completed by the participants including demographic and clinical characteristics. The primary efficacy outcomes were $HAMD_{17}$ and HAMA, both well-established interviewer-rated measures of depression and anxiety severity, respectively, from baseline to the end of treatment, and to the 1-month follow-up. For $HAMD_{17}$ [30], items are scored on a 5-point Likert scale (0-4) for a total score range from 0 to 68, with higher scores indicating more depressive symptoms. In the study sample, the total Cronbach alpha coefficient of HAMD17 was .89 (95% CI 0.86-0.92). Similarly, HAMA [31] consists of 14 items on a 5-point Likert scale (0-4) for a total score range from 0 to 56, with higher scores indicating more anxiety symptoms. Additionally, the internal consistency of this questionnaire was also good (α=.87, 95% CI 0.82-0.90). Given the severity of the COVID-19 epidemic, all assessors for $HAMD_{17}$ and HAMA scales from different centers were trained by one professional psychological assessment specialist through online conferences. The interviewer reliability is both high for $HAMD_{17}$ (intraclass correlation coefficient [ICC]=0.91, 95% CI 0.78-0.98) and HAMA (ICC=0.90, 95% CI 0.75-0.98).

Secondary end points for evaluating the efficacy of cCBT were SDS to assess self-rated depressive symptoms, SAS to assess anxiety symptoms, and AIS to assess insomnia symptoms. SDS [32] contains 20 items that reflect subjective feelings of depression. It is rated on a 4-point Likert scale (from 1, "no or a little of the time," to 4, "most of the time or all the time"), containing 10 symptom positive items and 10 symptom negative items. SAS [33] is also composed of 20 items and is rated on a 4-point Likert scale (from 1, "no or a little of the time," to 4, "most of the time or all the time"). Higher scores reflect more severe anxiety symptoms. The last secondary outcome measure is AIS [34], a validated brief questionnaire for a total score range from 0 to 24. Each item is measured on a 4-point Likert scale, with total scores between 4 and 6 representing suspicious symptoms of insomnia, scores higher than 6 representing insomnia, and scores less than 4 representing no insomnia. All three scales have been shown to have high internal consistency in this study sample (SDS: α=.93, 95% CI 0.90-0.96; SAS: α=.92, 95% CI 0.88-0.95; AIS: α=.87, 95% CI 0.81-0.92).

Sample Size

To estimate the sample size, a pilot study was conducted for measuring the $HAMD_{17}$ and HAMA score in 30 patients who received TAU. The standard deviation (σ) of the $HAMD_{17}$ and HAMA score in this group were both 10.50. Additionally, based on the previous studies on such interventions for depressive or anxiety symptoms [35-38], we expected to achieve a value δ≥5 in the level of depressive and anxiety symptoms with the cCBT intervention. Assuming a 5% significance level, two-tailed, and a power of 90%, a sample size of 188 patients (ie, 94 per group) was estimated. To include an estimated dropout rate of 25%, the entire sample size was increased from 188 to 250 participants (ie, 125 per group).

Statistical Analysis

SPSS software version 22.0 (IBM Corp) was used for all statistical analyses. Descriptive statistics were conducted for demographic and clinical characteristics in each treatment group. The measurement data were described by means and SDs. Two-sample *t* tests and chi-square or Fisher exact tests were used where appropriate, and these assessed demographic variable differences between the cCBT + TAU group and the TAU group. To confirm the improvements in each symptom at postintervention, the baseline and postintervention results of the dependent variables were analyzed using the paired *t* test, whereas the two-sample *t* test was used to detect differences between the treatment and control groups. Moreover, to explore the influence of gender, age, and education level on treatment response, we conducted a subgroup analysis to examine the changes in the scores of patients with different gender, age, and

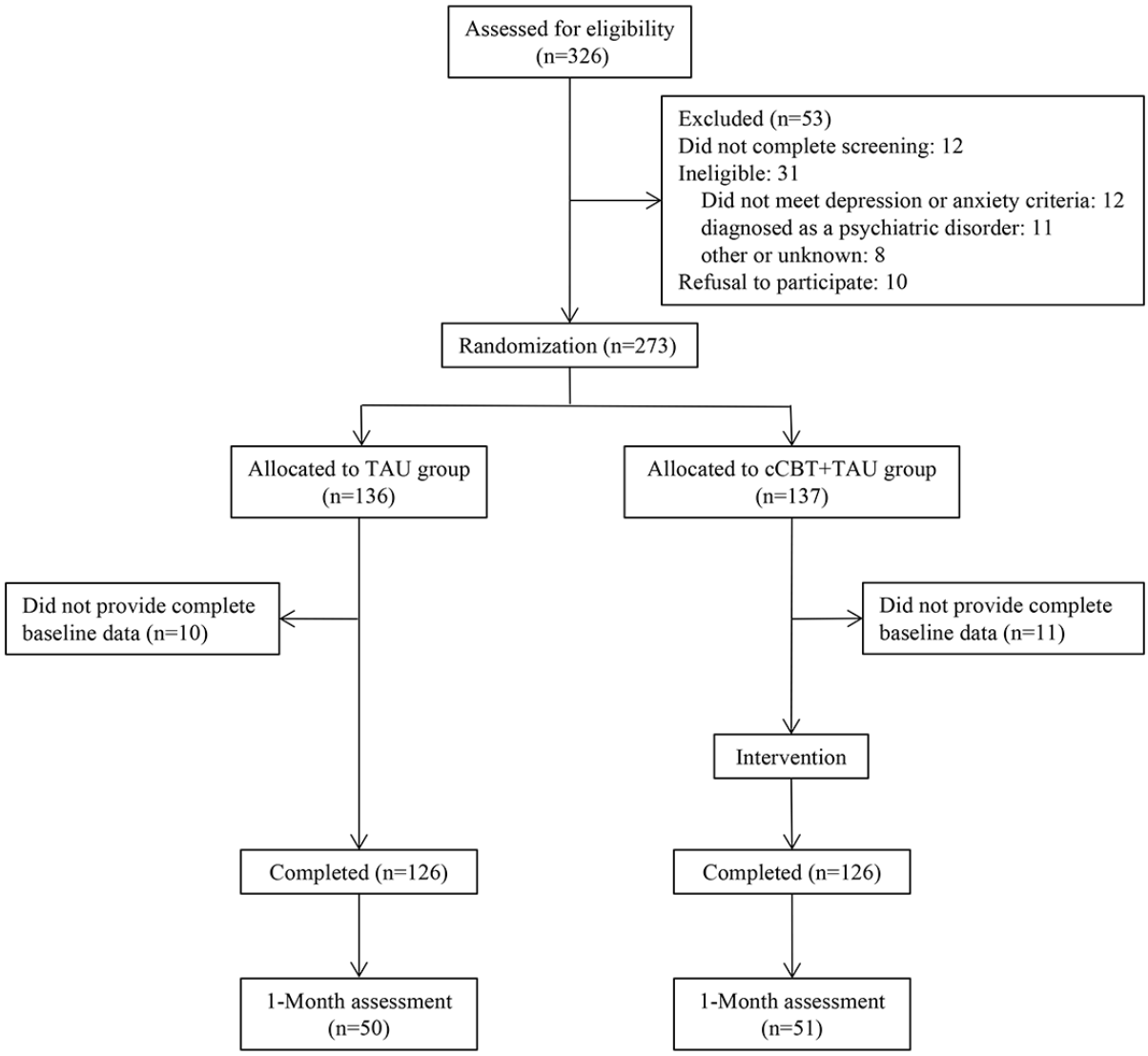
XSL•FO
RenderX

education level. For patients who finished the follow-up assessment, a mixed-effects model for repeated measures (MMRM) was used to compare data obtained at the baseline, postintervention, and follow-up periods between the cCBT + TAU group and the TAU group. Post hoc analysis was performed using the Bonferroni multiple comparison test. All tests were two-sided, and a *P* value less than .05 was considered statistically significant.

Results

A total of 326 patients with COVID-19 from five centers agreed to participate to the program. There were 273 participants that completed screening (see Figure 1 for details on reasons for exclusion); of these, 21 did not provide the complete baseline data and were therefore excluded. Thus, the analyzable sample consisted of the remaining 252 participants, who were randomly assigned to either the cCBT + TAU group (n=126) or the TAU group (n=126). All participants completed the assigned intervention (Figure 1). As shown in Table 1, no significant differences were found in the demographic characteristics between the cCBT + TAU group and the TAU group. None of the participants were taking any psychiatric medication at baseline and during the intervention. For physical symptoms of COVID-19, all patients were given corresponding treatments according to the Diagnosis and Treatment Protocol for COVID-19 issued by the National Health Commission, mainly including antiviral therapy, antibacterial therapy, and traditional Chinese medicine treatment. The results for the efficacy variables are presented in Table 2. Compared to the TAU group, the cCBT + TAU group showed significant improvement in both the primary outcomes of $HAMD_{17}$ and HAMA scores and the secondary outcomes of SDS, SAS, and AIS scores (all $P<.001$). Additionally, there were significant differences between the two groups in $HAMD_{17}$, HAMA, SDS, and SAS scores post intervention (all $P<.001$).

Figure 1. Flow diagram of the study. cCBT: computerized cognitive behavioral therapy; TAU: treatment as usual.



XSL•FO
RenderX

Table 1. Baseline characteristics of the full sample and follow-up sample.

Characteristics	cCBT[a] + TAU[b] group (n=126)	TAU group (n=126)	Difference (95% CI)	*t* test (*df*)	Chi-square (*df*)	*P* value	Cohen *d* or ϕ
The full sample							
Patients, n	126	126	N/A[c]	N/A	N/A	N/A	N/A
Sex, n			N/A	N/A	1.65 (1)	.20	0.08[d]
Male	70	80					
Female	56	46					
Age (years), mean (SD)	43.76 (14.31)	41.52 (11.51)	2.24 (–0.98 to 5.46)	1.37 (250)	N/A	.17	0.17
Education (years), mean (SD)	10.68 (3.91)	10.67 (4.39)	0.02 (–1.01 to 1.05)	0.03 (250)	N/A	.98	<0.01
Site, n			N/A	N/A	N/A	N/A	N/A
Affiliated Tongji Hospital of Huazhong University of Science and Technology	20	19					
Wuhan Institute for Tuberculosis Control	13	9					
Central Hospital of Chongqing Three Gorges	26	25					
Chongqing Public Health Medical Center	55	57					
The Fourth People's Hospital of Taiyuan	12	16					
Follow-up sample							
Patients, n	51	50	N/A	N/A	N/A	N/A	N/A
Sex, n			N/A	N/A	0.11 (1)	.74	0.03[d]
Male	31	32					
Female	20	18					
Age (years), mean (SD)	42.26 (12.66)	42.10 (10.58)	0.16 (–4.41 to 4.74)	0.07 (99)	N/A	.94	0.01
Education (years), mean (SD)	10.57 (4.21)	10.72 (4.79)	–0.15 (–1.92 to 1.61)	–0.17 (99)	N/A	.86	–0.03

[a]cCBT: computerized cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]N/A: not applicable.

[d]Indicates a ϕ value.

XSL•FO
RenderX

Table 2. Changes in the primary and secondary outcomes after the intervention between groups.

Measure	cCBT[a] + TAU[b] group (n=126)	TAU group (n=126)	Difference (95% CI)	*P* value[c]	Cohen *d*
$HAMD_{17}$[d]					
Baseline, mean (SD)	15.13 (3.33)	15.52 (3.43)	–0.39 (–1.23 to 0.45)	.36	–0.12
Postintervention, mean (SD)	8.19 (3.54)	15.20 (3.64)	–7.01(–7.90 to –6.12)	<.001	–1.95
Difference (95% CI)	6.94 (6.34 to 7.54)	0.32 (–0.04 to 0.68)	N/A[e]	N/A	N/A
P value[f]	<.001	.08	N/A	N/A	N/A
Cohen *d*	2.02	0.16	N/A	N/A	N/A
HAMA[g]					
Baseline, mean (SD)	14.52 (3.13)	13.97 (2.72)	0.56 (–0.17 to 1.28)	.13	0.18
Postintervention, mean (SD)	7.79 (3.60)	13.63 (3.24)	–5.84 (–6.69 to –4.99)	<.001	–1.70
Difference (95% CI)	6.73(6.13 to 7.33)	0.33 (–0.09 to 0.75)	N/A	N/A	N/A
P value	<.001	.12	N/A	N/A	N/A
Cohen *d*	1.97	0.14	N/A	N/A	N/A
SDS[h]					
Baseline, mean (SD)	45.89 (8.78)	45.66 (8.28)	0.23 (–1.89 to 2.35)	.83	0.03
Postintervention, mean (SD)	31.95 (6.88)	44.87 (7.48)	–12.92 (–14.70 to –11.14)	<.001	–1.79
Difference (95% CI)	13.94 (12.73 to 15.15)	0.79 (–0.19 to 1.76)	N/A	N/A	N/A
P value	<.001	.11	N/A	N/A	N/A
Cohen *d*	2.03	0.14	N/A	N/A	N/A
SAS[i]					
Baseline, mean (SD)	44.08 (10.44)	45.41 (7.71)	–1.33 (–3.61 to 0.95)	.25	–0.14
Postintervention, mean (SD)	30.37 (7.82)	44.53 (6.91)	–14.16 (–16.00 to –12.33)	<.001	–1.92
Difference (95% CI)	13.71 (12.11 to 15.30)	0.88 (–0.05 to 1.81)	N/A	N/A	N/A
P value	<.001	.06	N/A	N/A	N/A
Cohen *d*	1.51	0.40	N/A	N/A	N/A
AIS[j]					
Baseline, mean (SD)	8.98 (3.45)	8.67 (3.08)	0.31 (–0.50 to 1.12)	.45	0.09
Postintervention, mean (SD)	7.52 (2.99)	8.27 (3.22)	–0.75 (–1.52 to 0.03)	.05	–0.24
Difference (95% CI)	1.45 (1.04 to 1.86)	0.40 (–0.04 to 0.83)	N/A	N/A	N/A
P value	<.001	.07	N/A	N/A	N/A
Cohen *d*	0.63	0.16	N/A	N/A	N/A

[a]cCBT: computerized cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]*P* value of independent-samples *t* test.

[d]$HAMD_{17}$: Hamilton Depression Rating Scale.

[e]N/A: not applicable.

[f]*P* value of paired *t* test.

[g]HAMA: Hamilton Anxiety Scale.

[h]SDS: Self-Rating Depression Scale.

XSL•FO
RenderX

[i]SAS: Self-Rating Anxiety Scale.

[j]AIS: Athens Insomnia Scale.

Exploratory subgroup analyses examined the influence of gender, age, and education level on treatment response. For both male and female patients, the cCBT + TAU group showed some evidence for a greater reduction in $HAMD_{17}$, HAMA, SDS, and SAS scores compared with the TAU group. For female patients, we found no evidence of significant between-group differences for the change of AIS scores (Table S1 in Multimedia Appendix 1). Additionally, the cCBT + TAU group showed some evidence of greater improvement in symptoms of anxiety and depression compared with the TAU group, regardless of education level. However, changes in sleeping, as measured using AIS, did not differ between the cCBT + TAU group and TAU group for patients with middle school education (Table S2 in Multimedia Appendix 2). Additionally, the cCBT + TAU group also showed significantly decreased scores on all 5 scales compared with the TAU group, regardless of age (Table S3 in Multimedia Appendix 3).

Details about sociodemographic characteristics of participants who completed the 1-month follow-up assessment for each group are presented in Table 1. The results indicated that there were no significant differences between the cCBT + TAU group and TAU group for any of these variables. Besides, compared to the full sample in each condition, the follow-up sample showed no significant differences in demographic and psychological characteristics (Table 3). The patients who finished the follow-up assessment were analyzed using the MMRM and Bonferroni post hoc multiple comparison, with $HAMD_{17}$, HAMA, SDS, SAS, and AIS scores as the dependent variables, to reveal the relationship between the two groups and three periods.

Table 3. Comparison in demographic and psychological characteristics between full sample and follow-up sample.

Characteristics	cCBT[a] + TAU[b] group					TAU group				
	Full sample	Follow-up sample	Difference (95% CI)	*P* value	Cohen *d* or ϕ	Full sample	Follow-up sample	Difference (95% CI)	*P* value	Cohen *d* or ϕ
Sex, n			N/A[c]	.52	0.05[d]			N/A	.95	<0.01[d]
Male	70	31				80	31			
Female	56	20				46	18			
Age (years), mean (SD)	43.76 (14.31)	42.26 (12.66)	1.50 (–2.98 to 5.97)	.51	0.11	41.52 (11.51)	42.10 (10.58)	–0.58 (–4.29 to 3.14)	.76	–0.05
Education (years)	10.68 (3.91)	10.57 (4.21)	0.12 (–1.18 to 1.41)	.86	0.03	10.67 (4.39)	10.72 (4.79)	–0.05 (–1.54 to 1.43)	.94	–0.01
$HAMD_{17}$[e]	15.13 (3.33)	15.28 (2.23)	–0.15 (–1.16 to 0.81)	.73	–0.05	15.52 (3.43)	15.70 (1.28)	–0.18 (–1.17 to 0.80)	.71	–0.07
HAMA[f]	14.52 (3.13)	14.26 (2.31)	0.26 (–0.75 to 1.12)	.70	0.09	13.97 (2.72)	13.88 (1.86)	0.09 (–0.74 to 0.92)	.83	0.04
SDS[g]	45.89 (8.78)	46.10 (7.59)	–0.21 (–2.81 to 2.67)	.96	–0.03	45.66 (8.28)	45.22 (8.15)	0.44 (–2.28 to 3.19)	.75	0.05
SAS[h]	44.08 (10.44)	44.30 (10.49)	–0.22 (–3.13 to 3.66)	.88	–0.02	45.41 (7.71)	45.56 (7.25)	–0.15 (–2.65 to 2.35)	.91	–0.02
AIS[i]	8.98 (3.45)	8.58 (2.95)	0.41 (–0.82 to 1.34)	.64	0.12	8.67 (3.08)	8.20 (1.29)	0.47 (–0.42 to 1.36)	.30	0.20

[a]cCBT: computerized cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]N/A: not applicable.

[d]Indicates ϕ values.

[e]$HAMD_{17}$: Hamilton Depression Rating Scale.

[f]HAMA: Hamilton Anxiety Scale.

[g]SDS: Self-Rating Depression Scale.

[h]SAS: Self-Rating Anxiety Scale.

[i]AIS: Athens Insomnia Scale.

As demonstrated in Table 4, a significant difference in the interaction between groups and periods (all $P<.001$) was observed in all scale scores. Post hoc analysis revealed significant reduction in all five scores during the postintervention and follow-up period in the cCBT + TAU group from baseline (all $P<.05$, adjusted by the Bonferroni method).

XSL•FO
RenderX

Table 4. Repeated measures analysis of variance results for scores of $HAMD_{17}$, HAMA, SDS, SAS, and AIS between cCBT + TAU group and TAU group at baseline, postintervention, and follow-up.

Measure	cCBT[a] + TAU[b] group (n=51)	TAU group (n=50)	Difference (95% CI)	*P* value[c]	Partial η^2
$HAMD_{17}$[d]					
Baseline, mean (SD)	15.28 (2.23)	15.70 (1.28)	–0.42 (–1.22 to 0.38)	.30	0.02
Postintervention, mean (SD)	7.86 (3.04)[e]	15.46 (1.76)	–7.60 (–8.61 to –6.60)	<.001	0.82
Follow-up, mean (SD)	6.86 (1.77)[e,f]	15.26 (2.32)	–8.40 (–9.30 to –7.50)	<.001	0.87
P value	<.001[g]	<.001[h]	<.001[i]	N/A[j]	N/A
HAMA[k]					
Baseline, mean (SD)	14.26 (2.31)	13.88 (1.86)	0.38 (–0.41 to 1.17)	.34	0.02
Postintervention, mean (SD)	7.38 (2.84)[e]	13.24 (2.26)	–5.86 (–6.83 to –4.89)	<.001	0.75
Follow-up, mean (SD)	6.10 (2.04)[e,f]	13.20 (1.96)	–7.10 (–7.81 to –6.39)	<.001	0.89
P value	<.001[g]	<.001[h]	<.001[i]	N/A	N/A
SDS[l]					
Baseline, mean (SD)	46.10 (7.59)	45.22 (8.15)	0.88 (–2.48 to 4.24)	.60	0.01
Postintervention, mean (SD)	32.56 (6.54)[e]	45.56 (6.59)	–13.00 (–15.80 to –10.20)	<.001	0.64
Follow-up, mean (SD)	31.14 (5.65)[e,f]	44.70 (6.05)	–13.56 (–15.98 to –11.14)	<.001	0.72
P value	<.001[g]	<.001[h]	<.001[i]	N/A	N/A
SAS[m]					
Baseline, mean (SD)	44.30 (10.49)	45.56 (7.25)	–1.26 (–4.97 to 2.45)	.50	0.01
Postintervention, mean (SD)	29.66 (7.31)[e]	45.52 (7.05)	–15.86 (–18.97 to –12.75)	<.001	0.68
Follow-up, mean (SD)	29.12 (6.08)[e]	44.92 (5.92)	–15.80 (–18.34 to –13.27)	<.001	0.76
P value	<.001[g]	<.001[h]	<.001[i]	N/A	N/A
AIS[n]					
Baseline, mean (SD)	8.58 (2.95)	8.20 (1.29)	0.38 (–0.51 to 1.27)	.40	0.02
Postintervention, mean (SD)	6.98 (2.99)[e]	8.00 (2.22)	–1.02 (–2.11 to 0.07)	.06	0.07
Follow-up, mean (SD)	6.88 (2.72)[e]	7.82 (2.17)	–0.94 (–1.84 to –0.04)	.04	0.08
P value	<.001[g]	<.001[h]	.002[i]	N/A	N/A

[a]cCBT: computerized cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]*P* value adjusted by the Bonferroni method of between-group differences.

[d]$HAMD_{17}$: Hamilton Depression Rating Scale.

[e]Compared with baseline, $P<.05$, adjusted by the Bonferroni method.

[f]Compared with postintervention, $P<.05$, adjusted by the Bonferroni method.

[g]*P* value of group effect in the mixed-effects model for repeated measures.

[h]*P* value of time effect in the mixed-effects model for repeated measures.

[i]*P* value of interactive effects between time and group in the mixed-effects model for repeated measures.

[j]N/A: not applicable.

[k]HAMA: Hamilton Anxiety Scale.

[l]SDS: Self-Rating Depression Scale.

[m]SAS: Self-Rating Anxiety Scale.

[n]AIS: Athens Insomnia Scale.

XSL•FO
RenderX

Discussion

Principal Results

This study tested the effectiveness of the cCBT program we developed in patients with COVID-19 who have mild to moderate depressive or anxiety symptoms, compared with the TAU condition. The results implicated that the cCBT intervention was beneficial, given significant between-group differences in depression ($HAMD_{17}$, SDS), anxiety (HAMA, SAS), and insomnia (AIS). Other notable between-group observations were no significant differences in the improvement of insomnia among females and those with middle school education.

Facing public health emergencies, people experiencing stress responses are likely to encounter mental health problems [39]. The pandemic has increased the risk of psychological disorders, especially in confirmed cases of COVID-19 [40]. A meta-analysis about the prevalence of mental health problems in patients with COVID-19 showed prevalence rates of depression at 15.97%, anxiety at 15.5%, PTSD at 21.94%, and insomnia at 23.87% [41]—all far higher than usual [42]. Another survey of early rehabilitation on patients with COVID-19 showed that 43.1% of them had flashback symptoms, 27.8% had voiding symptoms, 40.5% had increased alertness, 22.2% had anxiety, and 38.1% had depression [43]. It can be implied that the main mental health problems of patients with COVID-19 were depression, anxiety, and PTSD. Selective serotonin reuptake inhibitor antidepressants are the first-line drug treatment for depression, anxiety, and PTSD [44,45]. Nevertheless, in the conditions of the pandemic, medication for COVID-19–prompted trauma responses is not the best option because it takes effect slowly. Moreover, only an approximate 33% of the patients were clinically cured [46], and these demonstrated residual symptoms following acute treatment [47]. Further, a review indicated there is a great potential for drug-to-drug interactions between antiretroviral drugs and psychotropics, especially antidepressants and anxiolytics, which could hinder treatment efficacy or even produce life-threatening adverse drug reactions [48].

As such, nonpharmacological treatment is considered an effective and safe option without adverse effects. CBT is the first-line psychological treatment recommended by the guidelines of depression, anxiety, and PTSD (ie, Canadian Network for Mood and Anxiety Treatments and National Institute of Health and Care Excellence guideline [49,50]. Li et al [51] designed a randomized controlled trial to prove traditional CBT significantly decreased depression, anxiety, and stress symptoms in patients with COVID-19 [51]. However, administering CBT effectively using face-to-face techniques is very time-consuming and costly, and is unavailable to most patients in isolation [52].

In this study, the independently developed cCBT program for patients with COVID-19 was used as a systematic cognitive, mood, and behavior intervention. Through this program, selected patients were subjected to 7 interventions for a period of 1 week, and each intervention was divided into three steps. Step 1 was designed to help patients acquire knowledge of novel coronavirus pneumonia, recognize their own stress statuses, learn the methods of adjusting negative emotions, educate themselves about postisolation, and practice behavioral relaxation training. Step 2 was designed using several games to reinforce the knowledge acquired from Step 1. Step 3 was designed to guide the patients in relaxation techniques.

The cCBT program we developed was used in five hospitals in China that isolated patients with COVID-19 during the outbreak. A total of 252 patients with COVID-19 were collected in this study and randomly divided into a cCBT + TAU group (n=126) and a TAU group (n=126); 51 patients in the cCBT + TAU group and 50 patients in the TAU group finished the follow-up assessment in one of the centers. We observed that the symptoms of depression, anxiety, and insomnia were relieved significantly in the patients who underwent the cCBT program. The results were consistent with other short-term studies about online or digital CBT treatment of mental health problems among patients with COVID-19 in the pandemic [53,54]. In the 1-month follow-up, the results of our study still showed a significant improvement of the efficacy in the cCBT + TAU group, which was in accordance with previous research on cCBT relieving depression, anxiety, and insomnia symptoms in the short-term follow-up [55,56]. However, some studies revealed that the efficacy of cCBT was no different from the control group after a long-term follow-up [57,58]. Mental status was influenced by various factors such as environment, social and family support, and psychological adjustment. Therefore, long-term follow-up needs to be further explored.

We subsequently explored subgroup analyses examining the influence of gender, age, and education level on treatment response. Compared to the TAU group, insomnia symptoms were not improved significantly in female patients and those with middle school education treated by cCBT. Previous studies indicated that females were a high-risk population for insomnia [59]. A meta-analysis revealed the prevalence of insomnia in females was higher than that in males [60]; also, females were the predicted risk factor of insomnia during the COVID-19 pandemic [61,62]. Our results supported previous studies. Zhang et al [63] investigated the relationship between insomnia and education levels in Chinese females, and they found there was a negative correlation between them. Another research study obtained the opposite result, which reported that higher education levels were associated with sleep disorders [62]. Notably, an investigation that focused on patients with COVID-19 who were treated in the Fangcang shelter hospital in Wuhan, China found that lower education levels were more likely to prompt depression, anxiety, and stress but had no impact on insomnia [64]. There are great differences in the results of previous research studies; the correlation between education levels and insomnia remains unclear, and our study provided a little evidence of that.

Advantages and Limitations

This study has several strengths. First, the cCBT program we independently developed is focused on patients with COVID-19 and is more suitable for the psychological adjustment of the patients than an ordinary cCBT program. Second, the use of a strict inclusion and exclusion standard ensured the homogeneity

XSL•FO
RenderX

of participants. We controlled patient inclusivity through scores on $HAMD_{17}$≥7 and <24 or HAMA≥7 and <21 to ensure that all patients had mild to moderate depressive and anxiety symptoms. The study involved five different hospitals from China to make the results representative. Finally, the patients were randomized to the cCBT + TAU group or the TAU group with a 1:1 ratio to balance confounding factors.

Inevitably, our trial has some limitations. First, participants in our trials were nonblinded, while the study looks at a multiplicity of outcomes, thereby increasing the risk for a type I error. We designed the trial initially to have patients and the evaluators blinded; however, to reduce the risk of infection, only a few doctors were permitted to enter the isolated ward. Such a situation impeded the use of blind methods in this study. Second, the sample sizes were relatively small, and the time before the follow-up was relatively short. In the future, we will continue to develop the cCBT program for a broader range of scenarios such as alleviating preoperative anxiety and depression in patients. Bigger sample sizes and a longer follow-up will be adopted to identify the long-term influence of cCBT, and the influencing factors of cCBT's efficacy will be analyzed in more detail. Additionally, we did not account for more confounding factors such as drug side effects and related adversity, which may be also highly relevant to the symptoms of depression and anxiety.

Conclusions

In sum, this was a prospective, multicenter, two-arm, parallel-group, and randomized controlled trial to identify the efficacy of a short-term cCBT program for treating symptoms of depression, anxiety, and insomnia among patients with COVID-19. The results suggested that the cCBT was significantly effective in relieving symptoms of depression, anxiety, and insomnia in patients with COVID-19 post intervention or after a 1-month follow-up. However, the insomnia symptoms in females and those with middle school education were more difficult to improve. Further research is needed to expand the sample size and to investigate the long-term effects of cCBT for symptoms of depression, anxiety, and insomnia.

Acknowledgments

We wish to express our thanks to all the patients for their participation. This study was supported by the National Natural Science Foundation of China (81601193), the Key Research and Development Project (International Cooperation) of Shanxi Province (201903D421059), the Transformation and Cultivation Project of Scientific and Technological Achievements of Universities in Shanxi Province ([2020]3), the Shanxi Province Science Foundation for Youths (2015021204), the Research Project Supported by Shanxi Scholarship Council of China (2015-100), and the First Hospital of Shanxi Medical University Foundation for Youths' Innovation (YC1409).

Authors' Contributions

ZL and Yong Xu contributed to the conception and design of the study. XL and DW administered the surveys. YY designed the program. XN, ST, YD, YJ, YW, and JZ collected the data. DQ and WZ analyzed the study data. DQ and Yifan Xu wrote the initial version of the manuscript. All authors contributed to the manuscript revision and both read and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Differences in the dependent variables after intervention between the treatment and control groups of males and females.
[DOC File , 68 KB - jmir_v23i5e26883_app1.doc]

Multimedia Appendix 2
Differences in the dependent variables after intervention between the treatment and control groups of different education levels.
[DOC File , 90 KB - jmir_v23i5e26883_app2.doc]

Multimedia Appendix 3
Differences in the dependent variables after intervention between the treatment and control groups of different ages.
[DOC File , 118 KB - jmir_v23i5e26883_app3.doc]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 8486 KB - jmir_v23i5e26883_app4.pdf]

References

XSL•FO
RenderX

1. Peiris JSM, Yuen KY, Osterhaus ADME, Stöhr K. The severe acute respiratory syndrome. N Engl J Med 2003 Dec 18;349(25):2431-2441. [doi: 10.1056/NEJMra032498] [Medline: 14681510]
2. Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. Lancet 2015 Sep 05;386(9997):995-1007 [FREE Full text] [doi: 10.1016/S0140-6736(15)60454-8] [Medline: 26049252]
3. Murray MJ. Ebola virus disease: a review of its past and present. Anesth Analg 2015 Sep;121(3):798-809. [doi: 10.1213/ANE.0000000000000866] [Medline: 26287303]
4. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]
5. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
6. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280. [doi: 10.1111/tmi.13383] [Medline: 32052514]
7. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-10-21]
8. 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. Lancet Psychiatry 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/S2215-0366(20)30046-8] [Medline: 32032543]
9. Romagnoli S, Peris A, De Gaudio AR, Geppetti P. SARS-CoV-2 and COVID-19: from the bench to the bedside. Physiol Rev 2020 Oct 01;100(4):1455-1466 [FREE Full text] [doi: 10.1152/physrev.00020.2020] [Medline: 32496872]
10. Rogers JP, Chesney E, Oliver D, Pollak TA, McGuire P, Fusar-Poli P, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. Lancet Psychiatry 2020 Jul;7(7):611-627 [FREE Full text] [doi: 10.1016/S2215-0366(20)30203-0] [Medline: 32437679]
11. Nie X, Wang Q, Wang M, Zhao S, Liu L, Zhu Y, et al. Anxiety and depression and its correlates in patients with coronavirus disease 2019 in Wuhan. Int J Psychiatry Clin Pract 2020 Jul 14:1-6. [doi: 10.1080/13651501.2020.1791345] [Medline: 32662692]
12. Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]
13. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the CoViD-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
14. Li W, Yang Y, Liu Z, Zhao Y, Zhang Q, Zhang L, et al. Progression of mental health services during the COVID-19 outbreak in China. Int J Biol Sci 2020;16(10):1732-1738 [FREE Full text] [doi: 10.7150/ijbs.45120] [Medline: 32226291]
15. Lam MH, Wing Y, Yu MW, Leung C, Ma RCW, Kong APS, et al. Mental morbidities and chronic fatigue in severe acute respiratory syndrome survivors: long-term follow-up. Arch Intern Med 2009 Dec 14;169(22):2142-2147. [doi: 10.1001/archinternmed.2009.384] [Medline: 20008700]
16. Bisson JI. Post-traumatic stress disorder. BMJ Clin Evid 2010 Feb 03;2010:1005 [FREE Full text] [Medline: 21718580]
17. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
18. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
19. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251. [doi: 10.1007/s10879-013-9243-y]
20. 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]
21. Motter JN, Pimontel MA, Rindskopf D, Devanand DP, Doraiswamy PM, Sneed JR. Computerized cognitive training and functional recovery in major depressive disorder: a meta-analysis. J Affect Disord 2016 Jan 01;189:184-191. [doi: 10.1016/j.jad.2015.09.022] [Medline: 26437233]
22. 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 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
23. Vallury KD, Jones M, Oosterbroek C. Computerized cognitive behavior therapy for anxiety and depression in rural areas: a systematic review. J Med Internet Res 2015 Jun 05;17(6):e139 [FREE Full text] [doi: 10.2196/jmir.4145] [Medline: 26048193]

XSL•FO
RenderX

24. 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]
25. Zhang SX, Wang Y, Jahanshahi AA, Li J, Schmitt VGH. Early evidence and predictors of mental distress of adults one month in the COVID-19 epidemic in Brazil. J Psychosom Res 2021 Mar;142:110366 [FREE Full text] [doi: 10.1016/j.jpsychores.2021.110366] [Medline: 33494004]
26. Broderick PC. Early adolescent gender differences in the use of ruminative and distracting coping strategies. J Early Adolescence 2016 Jul 26;18(2):173-191. [doi: 10.1177/0272431698018002003]
27. Li CE, DiGiuseppe R, Froh J. The roles of sex, gender, and coping in adolescent depression. Adolescence 2006;41(163):409-415. [Medline: 17225659]
28. Weissman MM. Treatment of depression: men and women are different? Am J Psychiatry 2014 Apr;171(4):384-387. [doi: 10.1176/appi.ajp.2013.13121668] [Medline: 24687191]
29. Dos Santos É, Molina ML, Mondin T, Cardoso TDA, Silva R, Souza L, et al. Long-term effectiveness of two models of brief psychotherapy for depression: a three-year follow-up randomized clinical trial. Psychiatry Res 2020 Apr;286:112804. [doi: 10.1016/j.psychres.2020.112804] [Medline: 32001003]
30. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23:56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
31. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
32. Dunstan DA, Scott N. Clarification of the cut-off score for Zung's self-rating depression scale. BMC Psychiatry 2019 Jun 11;19(1):177 [FREE Full text] [doi: 10.1186/s12888-019-2161-0] [Medline: 31185948]
33. Dunstan DA, Scott N. Norms for Zung's Self-rating Anxiety Scale. BMC Psychiatry 2020 Feb 28;20(1):90 [FREE Full text] [doi: 10.1186/s12888-019-2427-6] [Medline: 32111187]
34. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [doi: 10.1016/s0022-3999(00)00095-7] [Medline: 11033374]
35. Nakao S, Nakagawa A, Oguchi Y, Mitsuda D, Kato N, Nakagawa Y, et al. Web-based cognitive behavioral therapy blended with face-to-face sessions for major depression: randomized controlled trial. J Med Internet Res 2018 Sep 21;20(9):e10743 [FREE Full text] [doi: 10.2196/10743] [Medline: 30249583]
36. Ritvo P, Knyahnytska Y, Pirbaglou M, Wang W, Tomlinson G, Zhao H, et al. Online mindfulness-based cognitive behavioral therapy intervention for youth with major depressive disorders: randomized controlled trial. J Med Internet Res 2021 Mar 10;23(3):e24380 [FREE Full text] [doi: 10.2196/24380] [Medline: 33688840]
37. Rosso IM, Killgore WDS, Olson EA, Webb CA, Fukunaga R, Auerbach RP, et al. Internet-based cognitive behavior therapy for major depressive disorder: a randomized controlled trial. Depress Anxiety 2017 Mar;34(3):236-245 [FREE Full text] [doi: 10.1002/da.22590] [Medline: 28009467]
38. Yuan M, Ren Z, Zhu H, Zhang Y, Meng Y, Zhang W. [Regional homogeneity changes in patients with social anxiety disorders after cognitive behavioral therapy.]. Sichuan Da Xue Xue Bao Yi Xue Ban 2016 Nov;47(6):898-903. [Medline: 28598121]
39. Deng J, Zhou F, Hou W, Silver Z, Wong CY, Chang O, et al. The prevalence of depression, anxiety, and sleep disturbances in COVID-19 patients: a meta-analysis. Ann N Y Acad Sci 2021 Feb;1486(1):90-111 [FREE Full text] [doi: 10.1111/nyas.14506] [Medline: 33009668]
40. Khan S, Siddique R, Xiaoyan W, Zhang R, Nabi G, Sohail Afzal M, et al. Mental health consequences of infections by coronaviruses including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Brain Behav 2021 Feb;11(2):e01901. [doi: 10.1002/brb3.1901] [Medline: 33314732]
41. Cénat JM, Blais-Rochette C, Kokou-Kpolou CK, Noorishad P, Mukunzi JN, McIntee S, et al. Prevalence of symptoms of depression, anxiety, insomnia, posttraumatic stress disorder, and psychological distress among populations affected by the COVID-19 pandemic: A systematic review and meta-analysis. Psychiatry Res 2021 Jan;295:113599 [FREE Full text] [doi: 10.1016/j.psychres.2020.113599] [Medline: 33285346]
42. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
43. Yuan B, Cai X, Wang J, Hu X, Li Z, Lu T, Anyawen. Investigation on psychological stress state of COVID-19 patients in early rehabilitation and analysis of influencing factors. Chin J Clin Neurosurg 2020 Aug;25(08):531-534. [doi: 10.13798/j.issn.1009-153X.2020.08.011]
44. Walter HJ, Bukstein OG, Abright AR, Keable H, Ramtekkar U, Ripperger-Suhler J, et al. Clinical practice guideline for the assessment and treatment of children and adolescents with anxiety disorders. J Am Acad Child Adolesc Psychiatry 2020 Oct;59(10):1107-1124. [doi: 10.1016/j.jaac.2020.05.005] [Medline: 32439401]
45. Beaudoin CE. Post-traumatic stress disorder. In: Thompson TL, editor. Encyclopedia of Health Communication. Newbury Park, CA: SAGE Publications, Inc; 2014:1084-1086.

XSL•FO
RenderX

46. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-1917. [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]
47. Xiao L, Feng L, Zhu X, Wang G, Wu W, Hu Y. A national survey of residual symptoms in Chinese depressive patients after acute phase treatment. Chin J Psychiatry 2017;50(03):175-181. [doi: 10.3760/cma.j.issn.1006-7884.2017.03.005]
48. Mohebbi N, Talebi A, Moghadamnia M, Nazari Taloki Z, Shakiba A. Drug interactions of psychiatric and COVID-19 medications. Basic Clin Neurosci 2020;11(2):185-200 [FREE Full text] [doi: 10.32598/bcn.11.covid19.2500.1] [Medline: 32855778]
49. Parikh SV, Quilty LC, Ravitz P, Rosenbluth M, Pavlova B, Grigoriadis S, CANMAT Depression Work Group. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 2. Psychological Treatments. Can J Psychiatry 2016 Sep;61(9):524-539 [FREE Full text] [doi: 10.1177/0706743716659418] [Medline: 27486150]
50. Post-traumatic stress disorder. The National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng116 [accessed 2020-12-15]
51. Li J, Li X, Jiang J, Xu X, Wu J, Xu Y, et al. The effect of cognitive behavioral therapy on depression, anxiety, and stress in patients with COVID-19: a randomized controlled trial. Front Psychiatry 2020;11:580827. [doi: 10.3389/fpsyt.2020.580827] [Medline: 33192723]
52. Alavi N, Yang M, Stephenson C, Nikjoo N, Malakouti N, Layzell G, et al. Using the online psychotherapy tool to address mental health problems in the context of the COVID-19 pandemic: protocol for an electronically delivered cognitive behavioral therapy program. JMIR Res Protoc 2020 Dec 18;9(12):e24913 [FREE Full text] [doi: 10.2196/24913] [Medline: 33290245]
53. Wahlund T, Mataix-Cols D, Olofsdotter Lauri K, de Schipper E, Ljótsson B, Aspvall K, et al. Brief online cognitive behavioural intervention for dysfunctional worry related to the COVID-19 pandemic: a randomised controlled trial. Psychother Psychosom 2021;90(3):191-199 [FREE Full text] [doi: 10.1159/000512843] [Medline: 33212440]
54. Cheng P, Casement MD, Kalmbach DA, Castelan AC, Drake CL. Digital cognitive behavioral therapy for insomnia promotes later health resilience during the coronavirus disease 19 (COVID-19) pandemic. Sleep 2021 Apr 09;44(4):zsaa258 [FREE Full text] [doi: 10.1093/sleep/zsaa258] [Medline: 33249492]
55. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450. [doi: 10.1111/bjc.12055] [Medline: 24831119]
56. Sato D, Yoshinaga N, Nagai E, Nagai K, Shimizu E. Effectiveness of internet-delivered computerized cognitive behavioral therapy for patients with insomnia who remain symptomatic following pharmacotherapy: randomized controlled exploratory trial. J Med Internet Res 2019 Apr 11;21(4):e12686 [FREE Full text] [doi: 10.2196/12686] [Medline: 30973344]
57. Martínez V, Rojas G, Martínez P, Gaete J, Zitko P, Vöhringer PA, et al. Computer-assisted cognitive-behavioral therapy to treat adolescents with depression in primary health care centers in Santiago, Chile: a randomized controlled trial. Front Psychiatry 2019;10:552. [doi: 10.3389/fpsyt.2019.00552] [Medline: 31417440]
58. Qiu J, Chen W, Gao X, Xu Y, Tong H, Yang M, et al. A randomized controlled trial of group cognitive behavioral therapy for Chinese breast cancer patients with major depression. J Psychosom Obstet Gynaecol 2013 Jun;34(2):60-67. [doi: 10.3109/0167482X.2013.766791] [Medline: 23646866]
59. Guandalini LS, da Silva EF, Lopes JDL, Santos VB, Lopes CT, de Barros ALBL. Analysis of the evidence of related factors, associated conditions and at-risk populations of the NANDA-I nursing diagnosis insomnia. Int J Nurs Sci 2020 Oct 10;7(4):466-476 [FREE Full text] [doi: 10.1016/j.ijnss.2020.09.003] [Medline: 33195760]
60. Zeng L, Zong Q, Yang Y, Zhang L, Xiang Y, Ng CH, et al. Gender difference in the prevalence of insomnia: a meta-analysis of observational studies. Front Psychiatry 2020;11:577429. [doi: 10.3389/fpsyt.2020.577429] [Medline: 33329116]
61. Bajaj S, Blair KS, Schwartz A, Dobbertin M, Blair RJR. Worry and insomnia as risk factors for depression during initial stages of COVID-19 pandemic in India. PLoS One 2020;15(12):e0243527 [FREE Full text] [doi: 10.1371/journal.pone.0243527] [Medline: 33301508]
62. Fu W, Wang C, Zou L, Guo Y, Lu Z, Yan S, et al. Psychological health, sleep quality, and coping styles to stress facing the COVID-19 in Wuhan, China. Transl Psychiatry 2020 Jul 09;10(1):225. [doi: 10.1038/s41398-020-00913-3] [Medline: 32647160]
63. Zhang L, Ruan Y, Cui Y, Gu M. A prospective cohort study of insomnia factors in postmenopausal women. Chin J Clinician 2020;48(11):1349-1353. [doi: 10.3969/j.issn.2095-8552.2020.11.028]
64. Gu Y, Zhu Y, Xu F, Xi J, Xu G. Factors associated with mental health outcomes among patients with COVID-19 treated in the Fangcang shelter hospital in China. Asia Pac Psychiatry 2020 Nov 01:e12443. [doi: 10.1111/appy.12443] [Medline: 33135397]

Abbreviations

AIS: Athens Insomnia Scale

XSL•FO
RenderX

CBT: cognitive behavioral therapy
cCBT: computerized cognitive behavioral therapy
HAMA: Hamilton Anxiety Scale
$HAMD_{17}$: Hamilton Depression Rating Scale
ICC: intraclass correlation coefficient
MERS: Middle East respiratory syndrome
MMRM: mixed-effects model for repeated measures
PTSD: posttraumatic stress disorder
SARS: severe acute respiratory syndrome
SAS: Self-Rating Anxiety Scale
SDS: Self-Rating Depression Scale
TAU: treatment as usual

Edited by C Basch; submitted 01.01.21; peer-reviewed by M Casement, D Sato; comments to author 08.02.21; revised version received 14.03.21; accepted 24.04.21; published 14.05.21.

Please cite as:
Liu Z, Qiao D, Xu Y, Zhao W, Yang Y, Wen D, Li X, Nie X, Dong Y, Tang S, Jiang Y, Wang Y, Zhao J, Xu Y
The Efficacy of Computerized Cognitive Behavioral Therapy for Depressive and Anxiety Symptoms in Patients With COVID-19: Randomized Controlled Trial
J Med Internet Res 2021;23(5):e26883
URL: https://www.jmir.org/2021/5/e26883
doi:10.2196/26883
PMID:33900931

©Zhifen Liu, Dan Qiao, Yifan Xu, Wentao Zhao, Yang Yang, Dan Wen, Xinrong Li, Xiaoping Nie, Yongkang Dong, Shiyou Tang, Yi Jiang, Ying Wang, Juan Zhao, Yong Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.05.2021. 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

Factors Influencing Anxiety Among WeChat Users During the Early Stages of the COVID-19 Pandemic in Mainland China: Cross-sectional Survey Study

Changqing Zou[1*], MS; Weiyu Zhang[2*], BM; Kristin Sznajder[3], PhD; Fengzhi Yang[2], BM; Yajing Jia[2], BA; Ruqing Ma[2], BA; Can Cui[2], BA; Xiaoshi Yang[2], PhD

[1]Department of Humanities and Social Sciences, China Medical University, Shenyang, China

[2]Department of Social Medicine, School of Public Health, China Medical University, Shenyang, China

[3]Department of Public Health, College of Medicine, Pennsylvania State University, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Xiaoshi Yang, PhD
Department of Social Medicine, School of Public Health
China Medical University
No 77 Puhe Road, Shenyang North New Area
Liaoning Province
Shenyang, 110122
China
Phone: 86 18900910796
Email: xsyang@cmu.edu.cn

Abstract

Background: The rapid outbreak of COVID-19 around the world has adversely affected the mental health of the public. The prevalence of anxiety among the public has increased dramatically during the COVID-19 pandemic. However, there are few studies evaluating the effects of positive psychological responses and information-seeking behaviors on anxiety experienced among social media users during the COVID-19 pandemic.

Objective: This study evaluated the prevalence of anxiety and its associated factors among WeChat users in mainland China during the early stages of the COVID-19 pandemic.

Methods: From February 10 to February 24, 2020, a nationwide, web-based cross-sectional survey study was carried out using convenience sampling. Participants' levels of anxiety, positive psychological responses, and information-seeking behaviors were assessed. The survey was distributed among WeChat users via the WeChat smartphone platform. Chi-square tests and multivariable logistic regression analyses were performed to examine the factors associated with anxiety.

Results: This study found that the prevalence of anxiety (Generalized Anxiety Disorder 7-item [GAD-7] scale score ≥7) among WeChat users in China was 17.96% (446/2483) during the early stages of the COVID-19 pandemic. Results of multivariable logistic regression analysis showed that information-seeking behaviors such as cannot stop searching for information on COVID-19, being concerned about the COVID-19 pandemic, and spending more than 1 hour per day consuming information about the pandemic were found to be associated with increased levels of anxiety. Additionally, participants who chose social media and commercial media as the primary sources to obtain information about the COVID-19 pandemic were found more likely to report anxiety. Conversely, participants who were confident or rational about the COVID-19 pandemic were less likely to report anxiety.

Conclusions: This study found that positive psychological responses and information-seeking behaviors were closely associated with anxiety among WeChat users during the COVID-19 pandemic in China. It might be paramount to enhance mental well-being by helping people respond to the COVID-19 pandemic more rationally and positively in order to decrease symptoms of anxiety.

(J Med Internet Res 2021;23(5):e24412) doi:10.2196/24412

KEYWORDS

anxiety; COVID-19; information seeking behavior; positive psychological response; health information; public health emergency; mental health; online survey; China; cross-sectional study

XSL•FO
RenderX

Introduction

In January 2020, the World Health Organization classified COVID-19 as a public health emergency of international concern [1,2]. COVID-19 spread rapidly to 208 countries and regions and became a global pandemic, resulting in more than 30 million cases of infection and over 950,000 deaths worldwide, as of September 18, 2020 [3,4]. The sudden outbreak of COVID-19 adversely impacted the daily life and mental health of the public, which demanded urgent solutions [5-12].

The COVID-19 pandemic generated a psychological crisis among the public as the prevalence of mental disorders, including anxiety and depression, increased [13-18]. Various sources of information, including official media channels, social media, and the local government or community, could significantly affect the public's psychological responses to emergency events and their mental well-being. Recent research has found that excessive information-seeking behaviors have been closely related with mental disorders, including anxiety and depression, during the COVID-19 pandemic [19,20]. Individuals with anxiety tend to focus on negative information and correspondingly experience negative psychological responses [21]. The frequency and daily duration of social media use was higher in areas with a higher number of COVID-19 cases. High levels of social media consumption were associated with increased mental distress, especially among individuals with high levels of fear [22]. The metacognitive model illustrates that certain behaviors such as controlling behaviors, reassurance seeking, and checking behaviors are positively associated with anxiety [23-25]. Thus, excessive media consumption could lead to harmful psychological responses, which could increase anxiety among these individuals during the COVID-19 pandemic.

Positive psychological responses to the COVID-19 pandemic among the public play a crucial part in reducing anxiety. Positive thoughts and attitudes could help individuals cope with stressors [26]. A recent study found that hope can contribute to anxiety prevention [27]. The theory of rational emotive behavior therapy indicates that rational beliefs can alleviate symptoms of anxiety and other mental distress [28,29]. Additionally, cognitive behavioral models of health anxiety indicate that negative emotions and misinterpretations of health-related stimuli could increase the chances of developing anxiety [30-33]. Therefore, promoting positive psychological responses and initiating emotion regulation and positive perceptions of health-related information, such as keeping a positive attitude and being rational, are necessary for the public to be able to better cope with stress [34].

Researchers have studied the pathogenic and epidemiological characteristics, epidemiological trends, as well as the prevention and treatment of COVID-19 [35-43]. However, studies focused on the relationship between information-seeking behaviors and anxiety during the COVID-19 pandemic are sparse. Therefore, this study focused on the evaluation of the prevalence of anxiety and its association with positive psychological responses and information-seeking behaviors among users of WeChat—the most popular social media platform in Mainland China—during the COVID-19 pandemic.

Methods

Study Participants and Procedure

In order to examine the prevalence of anxiety among WeChat users during the COVID-19 pandemic, a cross-sectional study design using convenience sampling was employed via a web-based WeChat platform developed by the Department of Environmental Health of China Medical University in China. Due to the urgency of the COVID-19 pandemic and the necessity of timely acquisition, the questionnaire was first uploaded to this web-based WeChat platform. Next, the survey was distributed via all the open WeChat groups of the research assistants, to render the survey easily available to their close contacts. The survey was distributed between February 10 and February 24, 2020, during which we achieved a sample size that was adequate in accordance with previous related studies [44-47]. The questionnaire comprised the validated Chinese version of the Generalized Anxiety Disorder 7-item (GAD-7) scale, questions on positive psychological responses, and questions on information-seeking behaviors practiced during the early stage of the COVID-19 pandemic. The questionnaire took about 20 minutes to complete.

The inclusion criteria for participation in this study were as follows: at least 18 years of age, ability to read and write Chinese, ability to complete the web-based questionnaire by themselves using a smartphone, ability to offer electronic signed informed consent, and voluntary participation. The exclusion criteria were as follows: undergoing any therapy for psychological illness, history of any drug dependence, and diagnoses of any diseases or impairments that would prevent them from completing the questionnaire independently. The questionnaire had been preset for submission only after all the questions were answered within the range of the selected choices. Only data from complete questionnaires were analyzed.

Ethics Statement

This study was carried out in conformance with the Declaration of Helsinki (1989). The study protocols were approved by the Ethics Committee of China Medical University.

Demographic Characteristics of WeChat Users

The survey collected participants' demographic characteristics, including gender (male or female), age, marital status (married or other), occupation, education, and monthly income. Occupation was further classified into government worker, health care worker, professional staff (teacher, lawyer, journalist, etc), employee of enterprises, commercial personnel, soldier, student, and other. Education was further classified into junior college or below, bachelor's degree, and master's degree or above. Monthly income was categorized as follows: ≤¥5000 (≤US $725.19), ¥5001-10,000 (US $725.34-1450.39), and >¥10,000 (>US $1450.39).

Measurement of Anxiety

The validated Chinese version of the GAD-7, one of the most reliable tools to measure generalized anxiety disorder, was used

XSL•FO
RenderX

to assess anxiety among the study participants during the early stages of the COVID-19 pandemic [48]. Psychological problems related to anxiety were evaluated with a 4-point Likert-type scale, with options including "not at all sure=0", "several days=1", "over the half of the days=2" and "nearly every day=3," resulting in a total score ranging from 0 to 21. The cutoff score for anxiety was set at ≥7, based on the total GAD-7 score [49-51], with Cronbach α=.939 for GAD-7. The results of the factor analysis indicated that the seven items of the positive psychological responses tested were extracted as one component, which contributed to 73.518% of the variance (Kaiser-Meyer-Olkin [KMO]=0.935, *P*<.001]. The indicators of reliability and validity were good.

Measurement of Positive Psychological Responses

Positive psychological responses included "being confident," "being hopeful," and "being rational" in the past 2 weeks. They were measured by "yes" or "no" questions (Cronbach α=.788). The results of the factor analysis indicated that the three items of the positive psychological responses tested were extracted as one component, which contributed to 70.932% of the variance (KMO=.638, *P*<.001). The indicators of reliability and validity were good.

Measurement of Information-Seeking Behaviors

Information-seeking behaviors were assessed using three self-developed questions; these behaviors were assessed in the following four categories: (1) cannot stop searching for information about the COVID-19 pandemic, (2) being concerned about the COVID-19 pandemic, (3) time spent consuming information about the COVID-19 pandemic (ie, <1 h, 1-2 h, or ≥3 h), and (4) sources of information about the COVID-19 pandemic (social media and commercial media, central and local official media, basic-level government media, and community media). Responses to the item "cannot stop searching for information about the COVID-19 pandemic" included "agree," "not sure," and "disagree." Whether participants were concerned about the COVID-19 pandemic and their sources of information about the pandemic were assessed by "yes" or "no" questions.

Statistical Analyses

SPSS software (version 23.0; IBM Corp) was used to perform all statistical analyses. Chi-square tests were applied to assess bivariate associations with anxiety and the variables of interest. Anxiety was assessed using the binary variable (*anxiety* or *no anxiety*, as measured by the GAD-7 scale). Multivariable logistic regression analysis showed that the factors associated with anxiety were assessed while controlling for confounding variables. The variables included in the multivariable logistic regression analysis included age as a continuous variable and other variables as categorized variables. The responses were not included in the analysis when more than 95% of individuals had the same response to the categorical independent variables. Findings were considered statistically significant when a two-tailed *P* value was <.05.

Results

Demographic Characteristics and Prevalence of Anxiety Among WeChat Users

In this study, we observed that 446 of 2483 (17.96%) Chinese WeChat users who participated in our survey experienced anxiety. A total of 2501 adults participated in this survey and 2483 of them provided complete and logical answers, resulting in a valid response rate of 99.28%. The demographic characteristics of the participants and results from the bivariate analysis of anxiety are presented in Table 1. The mean age of participants was 34 (SD 12.82) years. Of the 2483 participants, 1550 (62.42%) were female, and 278 (17.94%) of them reported having experienced anxiety. About half of the participants (1239/2483, 49.90%) were married. Their occupations were classified into three groups according to their distribution and frequency: (1) government worker, health care worker, teacher, lawyer, or journalist (515/2483, 20.74%); (2) student (954/2483, 38.42%); and (3) other, including employee of enterprises, commercial personnel, or soldier (1014/2483, 40.84%). More than half of the participants had a bachelor's degree (1394/2483, 56.14%), and many had a master's degree or above (800/2483, 32.22%). The remaining 11.64% (289/2483) participants whose educational level was "some college or below" showed significantly higher prevalence of anxiety (61/289, 21.11%) than the other two groups. With regard to the distribution of monthly income, 40.03% (994/2483) of the participants earned ≤¥5000 (≤US $725.19), 35.60% (884/2483) earned ¥5001-10,000 (US $725.34-1450.39), and 24.37% (605/2483) earned >¥10,000 (>US $1450.39).

XSL•FO
RenderX

Table 1. Prevalence of anxiety and associated factors among study participants (N=2483).

Variable	Participants, n (%)	Anxiety experienced, n (%)	No anxiety experienced, n (%)	Chi-square (*df*)	*P* value
Demographic characteristics					
Gender				0.004 (1)	.96
Male	933 (37.60)	167 (17.90)	766 (82.10)		
Female	1550 (62.40)	278 (18.00)	1272 (82.00)		
Age (years)				1.985 (1)	.17
≤35	1474 (59.33)	278 (18.86)	1196 (81.14)		
>35	1009 (40.67)	168 (16.65)	841 (83.35)		
Marital status				0.071 (1)	.79
Married	1239 (49.90)	220 (17.76)	1019 (82.24)		
Other	1244 (50.10)	226 (18.17)	1018 (81.83)		
Occupation				4.042 (2)	.13
Government worker, health care worker, teacher, lawyer, journalist	515 (20.74)	85 (16.50)	430 (83.50)		
Student	954 (38.42)	160 (16.77)	794 (83.23)		
Other	1014 (40.84)	201 (19.82)	813 (80.18)		
Education				2.780 (2)	.25
College and below	289 (11.64)	61 (21.11)	228 (78.89)		
Bachelor's degree	1394 (56.14)	238 (17.07)	1156 (82.93)		
Master's degree and above	800 (32.22)	147 (18.38)	653 (81.63)		
Monthly income (¥)[a]				1.569 (2)	.46
≤5000	994 (40.03)	190 (19.11)	804 (80.89)		
5001-10,000	884 (35.60)	150 (16.97)	734 (83.03)		
>10,000	605 (24.37)	106 (17.52)	499 (82.48)		
Positive psychological responses					
Being confident				80.584 (1)	<.001
Yes	2339 (94.20)	380 (16.25)	1959 (83.75)		
No	144 (5.80)	66 (45.83)	78 (54.17)		
Being hopeful				47.537 (1)	<.001
Yes	2359 (95.01)	395 (16.74)	1964 (83.26)		
No	124 (4.99)	51 (41.13)	73 (58.87)		
Being rational				113.526 (1)	<.001
Yes	2281 (91.87)	354 (15.52)	1927 (84.48)		
No	202 (8.13)	92 (45.54)	110 (54.46)		
Information-seeking behaviors					
Cannot stop searching information about the COVID-19 pandemic				37.118 (1)	<.001
Yes	1474 (59.36)	322 (21.85)	1152 (78.15)		
No	1009 (40.64)	124 (12.29)	885 (87.71)		
Concerned about the COVID-19 pandemic				37.195 (1)	<.001
Yes	1223 (49.25)	278 (22.73)	945 (77.27)		
No	1260 (50.75)	168 (13.33)	1092 (86.67)		
Time spent consuming information about the COVID-19 pandemic (hours)				115.008 (2)	<.001

XSL•FO
RenderX

Variable	Participants, n (%)	Anxiety experienced, n (%)	No anxiety experienced, n (%)	Chi-square (*df*)	*P* value
<1	906 (36.48)	92 (10.15)	814 (89.85)		
1-2	1006 (40.52)	171 (17.00)	835 (83.00)		
≥3	571 (23.00)	183 (32.05)	388 (67.95)		
Sources of information about the COVID-19 pandemic					
Social media and commercial media				9.826 (1)	.001
Yes	2159 (86.95)	408 (18.90)	1751 (81.10)		
No	324 (13.05)	38 (11.73)	285 (87.96)		
Central official media				3.004 (1)	.09
Yes	2104 (84.74)	366 (17.40)	1738 (82.60)		
No	379 (15.26)	80 (21.11)	299 (78.89)		
Local official media, basic-level government and community				0.901 (1)	.36
Yes	1508 (60.73)	262 (17.37)	1246 (82.63)		
No	975 (39.27)	184 (18.87)	791 (81.13)		

[a]1¥=US $0.15

Positive Psychological Responses

The prevalence of anxiety according to the participants' positive psychological responses are presented in Table 1. We observed that the prevalence rate of anxiety was lower for the participants who felt confident (2339/2483, 94.20%), hopeful (2358/2483, 95.01%), or rational (2281/2483, 91.87%) about the COVID-19 pandemic (*P*<.001).

Information-Seeking Behaviors

The distribution and prevalence of anxiety with different information-seeking behaviors during the COVID-19 pandemic is shown in Table 1. In this study, participants who could not stop searching for information about the COVID-19 pandemic (1474/2483, 59.34%) or those who were concerned about the pandemic (1223/2483, 49.24%) were found to have a significantly higher prevalence of anxiety (*P*<.001) than other participants. The prevalence of anxiety was also observed to be higher for participants who spent more than 1 hour a day on searching for information about the COVID-19 pandemic (*P*<.001).

Our study findings also showed that participants who sought information about COVID-19 via social media and commercial media (2159/2483, 86.95%) had a higher prevalence of anxiety (408/2159, 18.90%) than those who did not seek information through social and commercial media (*P*=.001).

Factors Associated With Anxiety

Table 2 shows the final results of the multivariable logistic regression analysis. Participants who were confident (odds ratio [OR] 0.434, 95% CI 0.243-0.775) or rational (OR 0.286, 95% CI 0.202-0.405) about the COVID-19 pandemic were found to be less likely to experience anxiety. In contrast, the information-seeking behaviors among the study participants, including cannot stop searching for information on COVID-19 (OR 1.593, 95% CI 1.236-2.052) and being concerned about the COVID-19 pandemic (OR 1.389, 95% CI 1.080-1.788) were found to be associated with anxiety. Moreover, participants spending more than 1 hour a day consuming information about the COVID-19 pandemic were also found to be more likely to report anxiety (1-2 h: OR 1.622, 95% CI 1.209-2.176; ≥3 h: OR 3.915, 95% CI 2.823-5.430). Additionally, participants who chose social media and commercial media as the primary source of information about the COVID-19 pandemic were observed to be more likely to report anxiety (OR 1.531, 95% CI 1.043-2.246).

Thus, the final model of multivariable logistic regression and forest plot (Figure 1) indicated that the following factors were associated with anxiety: cannot stop searching for information about the COVID-19 pandemic, being concerned about the COVID-19 pandemic, spending more than 1 hour a day consuming information about the COVID-19 pandemic, and use of social media and commercial media as the primary source of information about the COVID-19 pandemic. Conversely, being confident and being rational were independently found to be inversely associated with anxiety.

XSL•FO
RenderX

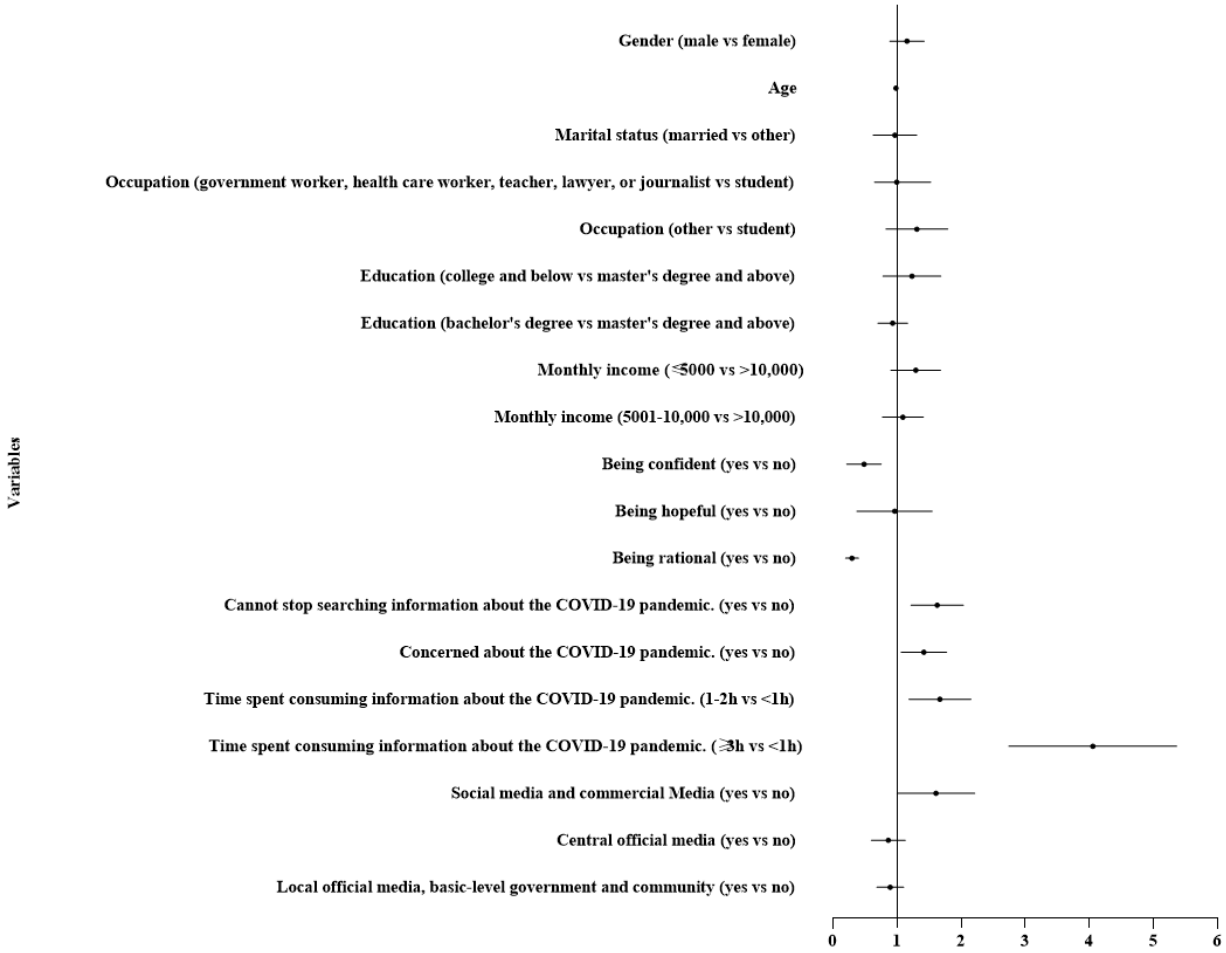
Table 2. Multivariable logistic regression analysis for factors associated with anxiety.

Variables	Odds ratio	95% CI
Demographic characteristics		
Gender (male vs female)	1.135	0.898-1.435
Age	0.981	0.967-0.995
Marital status (married vs other)	0.927	0.648-1.327
Occupation		
Government worker, health care worker, teacher, lawyer, journalist (vs student)	0.995	0.649-1.527
Other (vs student)	1.251	0.859-1.820
Education		
College and below (vs master's degree and above)	1.177	0.810-1.711
Bachelor's degree (vs master's degree and above)	0.913	0.710-1.174
Monthly income (¥)[a]		
≤5000 (vs >10,000)	1.253	0.925-1.699
5001-10,000 (vs >10,000)	1.061	0.788-1.427
Positive psychological responses		
Being confident (yes vs no)	0.434	0.243-0.775
Being hopeful (yes vs no)	0.843	0.444-1.602
Being rational (yes vs no)	0.286	0.202-0.405
Information-seeking behaviors		
Cannot stop searching information on COVID-19 (yes vs no)	1.593	1.236-2.052
Concerned about the COVID-19 pandemic (yes. vs no)	1.389	1.080-1.788
Time spent consuming information of COVID-19		
1-2 h (vs <1 h)	1.622	1.209-2.176
≥3 h (vs <1 h)	3.915	2.823-5.430
Sources of information about the COVID-19 pandemic		
Social media and commercial media (yes vs no)	1.531	1.043-2.246
Central official media (yes vs no)	0.836	0.611-1.143
Local official media, basic-level government, and community (yes vs no)	0.877	0.693-1.111

[a]1¥=US $0.15

XSL•FO
RenderX

Figure 1. Forest plot of factors associated with anxiety.



Discussion

Principal Findings

We observed that 17.96% of WeChat users in China who participated in this study reported having experienced anxiety during the early stages of the COVID-19 pandemic. This exceeds the prevalence of anxiety reported among the people in China before the pandemic (3.2%-8.1%) [52,53]. However, comparing the prevalence of anxiety during the severe acute respiratory syndrome (SARS) outbreak in 2002 (24.4%-35.0%), the prevalence of anxiety during the COVID-19 pandemic was slightly lower [54,55]. There remains a dearth of research on the relationship between information-seeking behaviors and anxiety during public health emergencies [56-58]. It is of vital importance to provide advice on how to cope with anxiety, including how to enhance positive psychological resources and prevent excessive information-seeking behaviors in order to prevent or reduce anxiety during public health emergencies. Our findings provide evidence-based recommendations on the promotion of positive psychological responses, psychological interventions, and information management to reduce anxiety for WeChat users in China.

Our study findings showed that anxiety was closely associated with positive psychological responses. Among the positive psychological responses, being confident and being rational were observed to be associated with lower prevalence of anxiety, which suggests that being confident and rational might prompt individuals to cope better with adverse events and difficult challenges during public health emergencies. Being rational also has been found to influence anxiety in previous studies, wherein people who were more positive coped better with life stressors and had low levels of anxiety [59,60]. WeChat users in this study had been exposed to a variety of stressors such as the perception of severe health risks of COVID-19, negative emotions associated with home quarantine, financial hardships caused by delays in work, and uncertainty during the pandemic, which might have resulted in reduced positive psychological responses and higher prevalence of anxiety. The framework of intolerance of uncertainty states that intolerance of uncertainty and cognitive avoidance are positively correlated with psychological responses and anxiety [23,61,62].

People could cope with stressors better and improve their mental health by means of rational-emotive behavioral interventions and cognitive behavioral therapies to reduce anxiety and improve psychological abilities, including emotion regulation and stress management [28,29,63,64]. The emotion dysregulation model of anxiety suggests that the lack of emotion regulation has a positive correlation with anxiety [65,66]. Therefore, to intensify protective effects of positive psychological responses, psychological interventions for the

XSL•FO
RenderX

public might be an effective approach, and these should be provided as soon as possible during public health emergencies.

This study also found that excessive consumption of information about COVID-19 might be closely linked with anxiety. Consuming information about the pandemic for more than 1 hour a day was observed to be associated with anxiety. Our results are consistent with other studies showing that anxiety is associated with increased use of smart phones, leading to excessive exposure to COVID-19 news [19,22,67,68]. The frequency and duration of searching for health information could exacerbate stress, anxiety, and perception of health risks [69-72]. WeChat users in this study might actively search or passively receive health information about the COVID-19 pandemic via multiple media sources. Searching for health information can be regarded as a source of anxiety and negative psychological responses. Spending too much time searching for health information could be linked to increased levels of anxiety, anger, or sadness. This might be due to increased fear of being infected by a severe disease and could adversely impact people's self-esteem and reduce their tolerance of uncertainty [73-81]. Conversely, it is also possible that anxious people are more likely to overconsume information. Health professionals should provide timely suggestions for the public on how to access and absorb health information from trusted sources [77,82,83].

Moreover, sources of information about COVID-19 were found to be significantly associated with anxiety, especially information obtained via social media was observed to have a positive correlation with anxiety. Excessively searching for health information on social media during traumatic events (eg, public health emergencies, food safety incidents, terror attacks, and natural disasters) can result in sleep disorders, distress such as anxiety and depression, and posttraumatic stress disorder [84-89]. Social media provides open platforms for the public to exchange their ideas and perspectives in a timely and prompt manner. However, it might be difficult for the public to distinguish true information from false information. Through improved dissemination of health information, coping strategies and healthy behaviors in the context of the COVID-19 pandemic could be encouraged, in addition to focusing on the prevention of negative psychological responses and improvement on positive psychological responses [90].

Limitations

Several limitations should be acknowledged. First, the data were collected using a self-administered web-based questionnaire via smartphones, which may have resulted in information bias and misclassification bias. It is possible that participants did not provide accurate information in the study in order to either be included in the study or to move quickly through the survey. To minimize the bias as much as possible, a pilot study was conducted to acquire participants' perception of the questionnaire. Therefore, the collected questionnaire had been filtered with data cleaning, checking for consistency and logicality of the answers, adjusting invalid and missing values. Due to the severity of the COVID-19 pandemic and the necessity of timely acquisition of data, the survey was distributed via WeChat rather than conducted using a face-to-face approach. Additionally, this survey among WeChat users used a convenience sampling method during the first 2 months of the COVID-19 pandemic. Therefore, generalizability of the results may be limited to the whole population of China. As the survey was shared within academic WeChat groups, it is possible the results reported in our study are from a population that has a higher education level than the general population, and this may have affected anxiety levels either positively or negatively, because the participants likely had a stronger understanding of the pandemic. Future studies should be carried out with random sampling. Finally, the results may be limited by residual confounding from unmeasured confounders such as regional economic disparity, allocation of health resources, and an individual's social status. To minimize the potential errors induced by unmeasured confounders, the survey was conducted across 3 municipalities and 22 provinces via WeChat, one of the most popular social platforms in China.

Conclusions

The COVID-19 pandemic has adversely affected the mental health of the public and resulted in a high prevalence of anxiety among them, especially during the early stages of the pandemic. Certain behaviors such as cannot stop searching for information on COVID-19, being concerned about the COVID-19 pandemic, spending more than 1 hour a day consuming information about the COVID-19 pandemic, and using social media and commercial media as the primary source of information about the COVID-19 pandemic were found to be associated with anxiety. In contrast, positive psychological responses such as being confident and being rational were found to be negatively correlated with anxiety. Prompt measures and guidance from public health authorities on choosing reliable and trusted sources of information might decrease the negative effects of overconsumption of COVID-19–related information and improve the overall mental health of the public.

Acknowledgments

This survey was conducted using a web-based WeChat platform developed by the Department of Environmental Health, China Medical University, which was approved by the Public Behavior Research Project of Prevention and Control of COVID-19, China Medical University (grant number: 121-1210120025). The researchers also appreciate all of the study participants for their time to complete the survey.

Authors' Contributions

WZ and CZ were responsible for data analysis and drafting the manuscript. KS was responsible for the revision of the manuscript. FY, YJ, RM, and CC were responsible for the explanation of the data. XY was responsible for the study design.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The multivariable logistic regression analysis of anxiety (without positive psychological responses).
[DOC File , 42 KB - jmir_v23i5e24412_app1.doc]

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
2. Archived: WHO Timeline - COVID-19. World Health Organization. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2021-04-29]
3. Coronavirus disease (COVID-19) Situation Report – 148. World Health Organization. URL: https://tinyurl.com/4pt69ybf [accessed 2021-04-29]
4. Real-time big data report of the epidemic [Data source: National Health Commission, People's Republic of China]. Webpage in Chinese. Baidu. URL: https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_3#tab0 [accessed 2021-04-29]
5. Dong L, Bouey J. Public mental health crisis during COVID-19 pandemic, China. Emerg Infect Dis 2020 Jul;26(7):1616-1618 [FREE Full text] [doi: 10.3201/eid2607.200407] [Medline: 32202993]
6. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Jun;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
7. Yang Y, Peng F, Wang R, Yange M, Guan K, Jiang T, et al. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun 2020 May;109:102434 [FREE Full text] [doi: 10.1016/j.jaut.2020.102434] [Medline: 32143990]
8. 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):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
9. Choi EPH, Hui BPH, Wan EYF. Depression and anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]
10. 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):3947 [FREE Full text] [doi: 10.3390/ijerph17113947] [Medline: 32498401]
11. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
12. Lima CKT, Carvalho PMM, Lima IAAS, Nunes JVAO, Saraiva JS, de Souza RI, et al. The emotional impact of Coronavirus 2019-nCoV (new Coronavirus disease). Psychiatry Res 2020 May;287:112915 [FREE Full text] [doi: 10.1016/j.psychres.2020.112915] [Medline: 32199182]
13. González-Sanguino C, Ausín B, Castellanos MA, Saiz J, López-Gómez A, Ugidos C, et al. Mental health consequences during the initial stage of the 2020 Coronavirus pandemic (COVID-19) in Spain. Brain Behav Immun 2020 Jul;87:172-176 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.040] [Medline: 32405150]
14. Asmundson GJG, Taylor S. How health anxiety influences responses to viral outbreaks like COVID-19: What all decision-makers, health authorities, and health care professionals need to know. J Anxiety Disord 2020 Apr;71:102211 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102211] [Medline: 32179380]
15. 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 Jan 21;94(3):210-214. [doi: 10.2471/blt.15.158543]
16. Ozamiz-Etxebarria N, Dosil-Santamaria M, Picaza-Gorrochategui M, Idoiaga-Mondragon N. Stress, anxiety, and depression levels in the initial stage of the COVID-19 outbreak in a population sample in the northern Spain. Webpage in Spanish. Cad Saude Publica 2020;36(4):e00054020 [FREE Full text] [doi: 10.1590/0102-311X00054020] [Medline: 32374806]
17. Huang Y, Zhao N. Mental health burden for the public affected by the COVID-19 outbreak in China: Who will be the high-risk group? Psychol Health Med 2021 Jan;26(1):23-34. [doi: 10.1080/13548506.2020.1754438] [Medline: 32286091]
18. Kola L. Global mental health and COVID-19. Lancet Psychiatry 2020 Aug;7(8):655-657 [FREE Full text] [doi: 10.1016/S2215-0366(20)30235-2] [Medline: 32502468]
19. Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, et al. Excessive media consumption about COVID-19 is associated with increased state anxiety: Outcomes of a large online survey in Russia. J Med Internet Res 2020 Sep 11;22(9):e20955 [FREE Full text] [doi: 10.2196/20955] [Medline: 32788143]

XSL•FO
RenderX

20. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
21. Herrera S, Montorio I, Cabrera I. Effect of anxiety on memory for emotional information in older adults. Aging Ment Health 2017 Apr;21(4):362-368. [doi: 10.1080/13607863.2015.1093601] [Medline: 26460518]
22. Bendau A, Petzold MB, Pyrkosch L, Mascarell Maricic L, Betzler F, Rogoll J, et al. Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. Eur Arch Psychiatry Clin Neurosci 2021 Mar;271(2):283-291 [FREE Full text] [doi: 10.1007/s00406-020-01171-6] [Medline: 32691135]
23. Behar E, DiMarco ID, Hekler EB, Mohlman J, Staples AM. Current theoretical models of generalized anxiety disorder (GAD): conceptual review and treatment implications. J Anxiety Disord 2009 Dec;23(8):1011-1023. [doi: 10.1016/j.janxdis.2009.07.006] [Medline: 19700258]
24. Wells A. A cognitive model of generalized anxiety disorder. Behav Modif 1999 Oct;23(4):526-555. [doi: 10.1177/0145445599234002] [Medline: 10533439]
25. James A, Wells A. Religion and mental health: towards a cognitive-behavioural framework. Br J Health Psychol 2003 Sep;8(Pt 3):359-376. [doi: 10.1348/135910703322370905] [Medline: 14606978]
26. Görgen SM, Hiller W, Witthöft M. Health anxiety, cognitive coping, and emotion regulation: a latent variable approach. Int J Behav Med 2014 Apr;21(2):364-374. [doi: 10.1007/s12529-013-9297-y] [Medline: 23436185]
27. Mirhosseini S, Dadgari A, Basirinezhad MH, Mohammadpourhodki R, Ebrahimi H. The role of hope to alleviate anxiety in COVID-19 outbreak among community dwellers: An online cross-sectional survey. Ann Acad Med Singap 2020 Oct;49(10):723-730 [FREE Full text] [Medline: 33283835]
28. David D, Cotet C, Matu S, Mogoase C, Stefan S. 50 years of rational-emotive and cognitive-behavioral therapy: A systematic review and meta-analysis. J Clin Psychol 2018 Mar;74(3):304-318 [FREE Full text] [doi: 10.1002/jclp.22514] [Medline: 28898411]
29. Eseadi C. Rational-emotive behavioral intervention helped patients with cancer and their caregivers to manage psychological distress and anxiety symptoms. World J Clin Oncol 2019 Mar 24;10(2):62-66 [FREE Full text] [doi: 10.5306/wjco.v10.i2.62] [Medline: 30815372]
30. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 2016 Dec;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
31. Schäfer JÖ, Naumann E, Holmes EA, Tuschen-Caffier B, Samson AC. Emotion regulation strategies in depressive and anxiety symptoms in youth: A meta-analytic review. J Youth Adolesc 2017 Feb;46(2):261-276. [doi: 10.1007/s10964-016-0585-0] [Medline: 27734198]
32. Gautreau CM, Sherry SB, Sherry DL, Birnie KA, Mackinnon SP, Stewart SH. Does catastrophizing of bodily sensations maintain health-related anxiety? A 14-day daily diary study with longitudinal follow-up. Behav Cogn Psychother 2015 Jul;43(4):502-512. [doi: 10.1017/S1352465814000150] [Medline: 24785012]
33. Hagger MS, Koch S, Chatzisarantis NLD, Orbell S. The common sense model of self-regulation: Meta-analysis and test of a process model. Psychol Bull 2017 Nov;143(11):1117-1154. [doi: 10.1037/bul0000118] [Medline: 28805401]
34. Grecucci A, Pappaianni E, Siugzdaite R, Theuninck A, Job R. Mindful emotion regulation: Exploring the neurocognitive mechanisms behind mindfulness. Biomed Res Int 2015;2015:670724 [FREE Full text] [doi: 10.1155/2015/670724] [Medline: 26137490]
35. Li H, Liu S, Yu X, Tang S, Tang C. Coronavirus disease 2019 (COVID-19): current status and future perspectives. Int J Antimicrob Agents 2020 May;55(5):105951 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105951] [Medline: 32234466]
36. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol 2020 Jul;146(1):110-118 [FREE Full text] [doi: 10.1016/j.jaci.2020.04.006] [Medline: 32294485]
37. 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]
38. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis 2020;34:101623 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101623] [Medline: 32179124]
39. Shi Y, Wang G, Cai X, Deng J, Zheng L, Zhu H, et al. An overview of COVID-19. J Zhejiang Univ Sci B 2020 May;21(5):343-360 [FREE Full text] [doi: 10.1631/jzus.B2000083] [Medline: 32425000]
40. Singhal T. A review of Coronavirus Disease-2019 (COVID-19). Indian J Pediatr 2020 Apr;87(4):281-286 [FREE Full text] [doi: 10.1007/s12098-020-03263-6] [Medline: 32166607]
41. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]

XSL•FO
RenderX

42. Wang J, Pan L, Tang S, Ji JS, Shi X. Mask use during COVID-19: A risk adjusted strategy. Environ Pollut 2020 Nov;266(Pt 1):115099 [FREE Full text] [doi: 10.1016/j.envpol.2020.115099] [Medline: 32623270]
43. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol 2020 Jun;92(6):568-576 [FREE Full text] [doi: 10.1002/jmv.25748] [Medline: 32134116]
44. Hu G, Rao K, Hu M, Sur Z. Preparing for and responding to public health emergencies in China: a focus group study. J Public Health Policy 2007 Jul;28(2):185-195. [doi: 10.1057/palgrave.jphp.3200130] [Medline: 17585319]
45. 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]
46. Kamangar F, Islami F. Sample size calculation for epidemiologic studies: principles and methods. Arch Iran Med 2013 May;16(5):295-300 [FREE Full text] [Medline: 23641744]
47. Liu Y, Guo N, Li T, Zhuang W, Jiang H. Prevalence and associated factors of postpartum anxiety and depression symptoms among women in Shanghai, China. J Affect Disord 2020 Sep 01;274:848-856. [doi: 10.1016/j.jad.2020.05.028] [Medline: 32664024]
48. Spitzer RL, Kroenke K, Williams JBW, 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]
49. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
50. Budikayanti A, Larasari A, Malik K, Syeban Z, Indrawati LA, Octaviana F. Screening of generalized anxiety disorder in patients with epilepsy: Using a valid and reliable Indonesian version of Generalized Anxiety Disorder-7 (GAD-7). Neurol Res Int 2019;2019:5902610 [FREE Full text] [doi: 10.1155/2019/5902610] [Medline: 31275648]
51. Zhong Q, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sánchez SE, et al. Diagnostic validity of the Generalized Anxiety Disorder - 7 (GAD-7) among pregnant women. PLoS One 2015;10(4):e0125096 [FREE Full text] [doi: 10.1371/journal.pone.0125096] [Medline: 25915929]
52. Hu Q, Wan Y, Su L, Li H, Jin Y, Li T. Prevalence of anxiety disorder among mainland residents in China: A meta-analysis. Article in Chinese. Chin J Psychiatry 2013;46(4):204-211. [doi: 10.3760/cma.j.issn.1006-7884.2013.04.004]
53. Tao J, Gong F, Lu J, Liang L, Liao X. Meta-analysis on the prevalence of depressive and /or anxiety disorder in general hospitals in China. Article in Chinese. Sichuan Mental Health 2018;31:73-78. [doi: 10.11886/j.issn.1007-3256.2018.01.018]
54. Fu C, Chen L, Lei L. Investigation and analysis of anxiety status of college students during the period of SARS. Article in Chinese. J of Health Psychol 2003;11(6):434-436. [doi: 10.13342/j.cnki.cjhp.2003.06.015]
55. Cheng SKW, Wong CW, Tsang J, Wong KC. Psychological distress and negative appraisals in survivors of severe acute respiratory syndrome (SARS). Psychol Med 2004 Oct;34(7):1187-1195. [doi: 10.1017/s0033291704002272] [Medline: 15697045]
56. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
57. Que J, Shi L, Deng J, Liu J, Zhang L, Wu S, et al. Psychological impact of the COVID-19 pandemic on healthcare workers: a cross-sectional study in China. Gen Psychiatr 2020;33(3):e100259 [FREE Full text] [doi: 10.1136/gpsych-2020-100259] [Medline: 32596640]
58. Wei N, Huang B, Lu S, Hu J, Zhou X, Hu C, et al. Efficacy of internet-based integrated intervention on depression and anxiety symptoms in patients with COVID-19. J Zhejiang Univ Sci B 2020 May;21(5):400-404 [FREE Full text] [doi: 10.1631/jzus.B2010013] [Medline: 32425006]
59. Carretta CM, Ridner SH, Dietrich MS. Hope, hopelessness, and anxiety: a pilot instrument comparison study. Arch Psychiatr Nurs 2014 Aug;28(4):230-234. [doi: 10.1016/j.apnu.2014.05.005] [Medline: 25017555]
60. Germann JN, Leonard D, Heath CL, Stewart SM, Leavey PJ. Hope as a predictor of anxiety and depressive symptoms following pediatric cancer diagnosis. J Pediatr Psychol 2018 Mar 01;43(2):152-161. [doi: 10.1093/jpepsy/jsx097] [Medline: 29049751]
61. Dugas MJ, Marchand A, Ladouceur R. Further validation of a cognitive-behavioral model of generalized anxiety disorder: diagnostic and symptom specificity. J Anxiety Disord 2005;19(3):329-343. [doi: 10.1016/j.janxdis.2004.02.002] [Medline: 15686860]
62. Dunsmoor JE, Paz R. Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. Biol Psychiatry 2015 Sep 01;78(5):336-343. [doi: 10.1016/j.biopsych.2015.04.010] [Medline: 25981173]
63. Noormohamadi SM, Arefi M, Afshaini K, Kakabaraee K. The effect of rational-emotive behavior therapy on anxiety and resilience in students. Int J Adolesc Med Health Epub ahead of print posted online on June 8, 2019. [doi: 10.1515/ijamh-2019-0099] [Medline: 31188779]

XSL•FO
RenderX

64. Sahranavard S, Esmaeili A, Dastjerdi R, Salehiniya H. The effectiveness of stress-management-based cognitive-behavioral treatments on anxiety sensitivity, positive and negative affect and hope. Biomedicine (Taipei) 2018 Dec;8(4):23 [FREE Full text] [doi: 10.1051/bmdcn/2018080423] [Medline: 30474604]
65. Kirwan M, Pickett SM, Jarrett NL. Emotion regulation as a moderator between anxiety symptoms and insomnia symptom severity. Psychiatry Res 2017 Aug;254:40-47. [doi: 10.1016/j.psychres.2017.04.028] [Medline: 28448803]
66. Suveg C, Morelen D, Brewer GA, Thomassin K. The Emotion Dysregulation Model of Anxiety: a preliminary path analytic examination. J Anxiety Disord 2010 Dec;24(8):924-930. [doi: 10.1016/j.janxdis.2010.06.018] [Medline: 20634040]
67. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
68. 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]
69. Muse K, McManus F, Leung C, Meghreblian B, Williams JMG. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. J Anxiety Disord 2012 Jan;26(1):189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]
70. 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]
71. Sell TK, Watson C, Meyer D, Kronk M, Ravi S, Pechta LE, et al. Frequency of risk-related news media messages in 2016 coverage of Zika virus. Risk Anal 2018 Dec;38(12):2514-2524. [doi: 10.1111/risa.12961] [Medline: 29314118]
72. Elhai JD, Hall BJ, Erwin MC. Emotion regulation's relationships with depression, anxiety and stress due to imagined smartphone and social media loss. Psychiatry Res 2018 Mar;261:28-34. [doi: 10.1016/j.psychres.2017.12.045] [Medline: 29276991]
73. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
74. Stone J, Sharpe M. Internet resources for psychiatry and neuropsychiatry. J Neurol Neurosurg Psychiatry 2003 Jan;74(1):10-12 [FREE Full text] [doi: 10.1136/jnnp.74.1.10] [Medline: 12486258]
75. Gureje O, Ustün TB, Simon GE. The syndrome of hypochondriasis: a cross-national study in primary care. Psychol Med 1997 Sep;27(5):1001-1010. [doi: 10.1017/s0033291797005345] [Medline: 9300506]
76. Norr AM, Capron DW, Schmidt NB. Medical information seeking: impact on risk for anxiety psychopathology. J Behav Ther Exp Psychiatry 2014 Sep;45(3):402-407. [doi: 10.1016/j.jbtep.2014.04.003] [Medline: 24818986]
77. Sell TK, Boddie C, McGinty EE, Pollack K, Smith KC, Burke TA, et al. News media coverage of U.S. Ebola policies: Implications for communication during future infectious disease threats. Prev Med 2016 Dec;93:115-120. [doi: 10.1016/j.ypmed.2016.09.016] [Medline: 27664539]
78. Wormwood JB, Devlin M, Lin Y, Barrett LF, Quigley KS. When words hurt: Affective word use in daily news coverage impacts mental health. Front Psychol 2018;9:1333 [FREE Full text] [doi: 10.3389/fpsyg.2018.01333] [Medline: 30116210]
79. 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;25(3):649-660 [FREE Full text] [doi: 10.1177/1460458217719561] [Medline: 28728457]
80. 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]
81. Apaolaza V, Hartmann P, D'Souza C, Gilsanz A. Mindfulness, compulsive mobile social media use, and derived stress: The mediating roles of self-esteem and social anxiety. Cyberpsychol Behav Soc Netw 2019 Jun;22(6):388-396. [doi: 10.1089/cyber.2018.0681] [Medline: 31070455]
82. Gardner B, Smith L, Mansfield L. How did the public respond to the 2015 expert consensus public health guidance statement on workplace sedentary behaviour? A qualitative analysis. BMC Public Health 2017 Feb 02;17(1):47 [FREE Full text] [doi: 10.1186/s12889-016-3974-0] [Medline: 28148236]
83. Caburnay CA, Kreuter MW, Luke DA, Logan RA, Jacobsen HA, Reddy VC, et al. The news on health behavior: coverage of diet, activity, and tobacco in local newspapers. Health Educ Behav 2003 Dec;30(6):709-722. [doi: 10.1177/1090198103255456] [Medline: 14655865]
84. Goodwin R, Lemola S, Ben-Ezra M. Media use and insomnia after terror attacks in France. J Psychiatr Res 2018 Mar;98:47-50. [doi: 10.1016/j.jpsychires.2017.12.006] [Medline: 29276963]
85. Leroy F, Brengman M, Ryckbosch W, Scholliers P. Meat in the post-truth era: Mass media discourses on health and disease in the attention economy. Appetite 2018 Jun 01;125:345-355. [doi: 10.1016/j.appet.2018.02.028] [Medline: 29486208]
86. Oh HJ, Lee H. When do people verify and share health rumors on social media? The effects of message importance, health anxiety, and health literacy. J Health Commun 2019;24(11):837-847. [doi: 10.1080/10810730.2019.1677824] [Medline: 31609678]

XSL•FO
RenderX

87. Fukasawa M, Kawakami N, Nakayama C, Yasumura S. Relationship between use of media and radiation anxiety among the residents of Fukushima 5.5 years after the nuclear power plant accident. Disaster Med Public Health Prep 2021 Mar;15(1):42-49. [doi: 10.1017/dmp.2019.132] [Medline: 31779727]
88. Orui M, Nakayama C, Kuroda Y, Moriyama N, Iwasa H, Horiuchi T, et al. The association between utilization of media information and current health anxiety among the Fukushima Daiichi nuclear disaster evacuees. Int J Environ Res Public Health 2020 Jun 01;17(11):3921 [FREE Full text] [doi: 10.3390/ijerph17113921] [Medline: 32492886]
89. Pandey A, Patni N, Singh M, Sood A, Singh G. YouTube as a source of information on the H1N1 influenza pandemic. Am J Prev Med 2010 Mar;38(3):e1-e3. [doi: 10.1016/j.amepre.2009.11.007] [Medline: 20171526]
90. Basch CH, Hillyer GC, Erwin ZM, Mohlman J, Cosgrove A, Quinones N. News coverage of the COVID-19 pandemic: Missed opportunities to promote health sustaining behaviors. Infect Dis Health 2020 Aug;25(3):205-209 [FREE Full text] [doi: 10.1016/j.idh.2020.05.001] [Medline: 32426559]

Abbreviations

GAD-7: Generalized Anxiety Disorder 7-item
KMO: Kaiser-Meyer-Olkin
OR: odds ratio
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 18.09.20; peer-reviewed by B Sousa-Pinto, T Burke; comments to author 11.10.20; revised version received 28.10.20; accepted 17.04.21; published 17.05.21.

Please cite as:
Zou C, Zhang W, Sznajder K, Yang F, Jia Y, Ma R, Cui C, Yang X
Factors Influencing Anxiety Among WeChat Users During the Early Stages of the COVID-19 Pandemic in Mainland China: Cross-sectional Survey Study
J Med Internet Res 2021;23(5):e24412
URL: https://www.jmir.org/2021/5/e24412
doi:10.2196/24412
PMID:33878025

©Changqing Zou, Weiyu Zhang, Kristin Sznajder, Fengzhi Yang, Yajing Jia, Ruqing Ma, Can Cui, Xiaoshi Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2021. 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

SARS-CoV-2: The Second Wave in Europe

Athanassios S Fokas[1,2,3], MD, PhD; George A Kastis[3], PhD

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom
[2]Viterbi School of Engineering, University of South California, Los Angeles, CA, United States
[3]Mathematics Research Center, Academy of Athens, Athens, Greece

Corresponding Author:
Athanassios S Fokas, MD, PhD
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Wilberforce Rd
Cambridge, CB3 0WA
United Kingdom
Phone: 44 1223 339
Email: tf227@cam.ac.uk

Abstract

Although the SARS-CoV-2 virus has already undergone several mutations, the impact of these mutations on its infectivity and virulence remains controversial. In this viewpoint, we present arguments suggesting that SARS-CoV-2 mutants responsible for the second wave have less virulence but much higher infectivity. This suggestion is based on the results of the forecasting and mechanistic models developed by our study group. In particular, in May 2020, the analysis of our mechanistic model predicted that the easing of lockdown measures will lead to a dramatic second wave of the COVID-19 outbreak. However, after the lockdown was lifted in many European countries, the resulting number of reported infected cases and especially the number of deaths remained low for approximately two months. This raised the false hope that a substantial second wave will be avoided and that the COVID-19 epidemic in these European countries was nearing an end. Unfortunately, since the first week of August 2020, the number of reported infected cases increased dramatically. Furthermore, this was accompanied by an increasingly large number of deaths. The rate of reported infected cases in the second wave was much higher than that in the first wave, whereas the rate of deaths was lower. This trend is consistent with higher infectivity and lower virulence. Even if the mutated form of SARS-CoV-2 is less virulent, the very high number of reported infected cases implies that a large number of people will perish.

(J Med Internet Res 2021;23(5):e22431) doi:10.2196/22431

KEYWORDS

mathematical modelling of epidemics; COVID-19; SARS CoV-2; pandemic; lockdown in Europe

Introduction

The novel coronavirus SARS-CoV-2 has already undergone several mutations. A study claimed that a specific mutation had created a more “aggressive” form of the virus [1]. However, it was later argued that, during epidemics in general, a virus spreads to new areas and creates new localized epidemics that grow exponentially [2]. This finding implies that a mutated virus spreading in these new uninfected areas will rapidly increase its mutation frequency; this will occur even if the mutation is neutral or detrimental to the virus itself. This study also showed that the observations reported by Tang et al [1] could be explained by the abovementioned process and that there was *no* evidence for the emergence of a more aggressive form of the virus [2]. Another study analyzed a set of mutations in the spike protein of SARS-CoV-2 and concluded that a specific mutation had enhanced the capacity of the virus to spread from China to Europe, North America, and Australia [3]. These authors concluded that this was the result of the fact that this mutation had made the virus more “transmissible,” allowing it to outcompete other virus strains that did not possess this mutation. However, this conclusion was challenged by several scientists. In particular, a more formal analysis of the effect of this mutation by another research team suggested that it actually has reduced transmissibility [4]. It has now been firmly established that the second wave was caused by a new mutated form of SARS-CoV-2, called D614G—the name refers to the substitution of the amino acid aspartic acid (D) by glycine (G) in a region of the viral genome that encodes the relevant spike protein. In particular, a study conducted in Houston, Texas, United States, found that D614G is the fully dominant form of the second wave of the COVID-19 pandemic, exhibiting a

XSL•FO
RenderX

prevalence of 99.9% [5]. The study also found that people infected with the new SARS-CoV-2 strain had higher loads of the virus in their upper respiratory tracts, which suggests that the D614G strain may be more infectious; however, no definitive conclusion regarding its infectivity has been reached. The high prevalence of the new strain was also established in a study in the United Kingdom involving the analysis of 25,000 whole-genome SARS-CoV-2 sequences [6], but no conclusions were reached regarding its infectivity or virulence. Therefore, the crucial question of establishing whether the new strain D614G was indeed more infectious and/or less virulent remained open [7].

In the framework introduced by Holmdahl and Buckee [8], epidemiological models are broadly divided into two categories: *forecasting* and *mechanistic*. The former models fit a specific curve to the data and then attempt to predict the dynamics of the quantity under consideration. The main limitation of these models is that they remain valid only if the epidemiological situation remains unchanged. For example, they can be used during a lockdown period but will not make accurate predictions after the lockdown is lifted. In contrast to forecasting models, mechanistic models can make predictions even when the relevant circumstances change. Their main limitation is the difficulty of determining the parameters specifying such models.

Our group has developed both forecasting and mechanistic models. In this paper, we present the implications of our models to argue that the mutated virus D614G was more infectious but less virulent.

The Prediction of the Effect of Easing Lockdown Measures by Using a Mechanistic Type Model

The generic weakness of mechanistic models, namely the difficulty in determining the parameters of the given model, was partially bypassed in a previous study [9]. Indeed, this study showed that the death data *uniquely* determine the parameters specifying a given ordinary differential equation characterizing the number of deaths. Furthermore, a robust numerical algorithm was presented for obtaining these parameters. The abovementioned equation was obtained from the manipulation of a susceptible-exposed-infected-recovered (SEIR) type model. Using this methodology, predictions were made about the dynamics of the COVID-19 epidemics in Greece, Andalusia, and Portugal, following the easing of lockdown measures. One of the above parameters obtained from the death data during the lockdown period corresponds to the number of contacts of the asymptomatic individuals. By changing this parameter (doubling, tripling, etc), while keeping all other parameters unchanged, the effect of relaxing the lockdown measures could be *quantified*: if the contacts of the asymptomatic people double, nothing much happens; but if they triple, then the situation dramatically deteriorates. The numbers of reported infected cases during the second wave was within the range of our prediction made in May 2020; fortunately, the number of deaths reported was lower, suggesting less virulence.

Cumulative Number of Deaths and Cumulative Number of Reported Infected Cases

It is correctly noted in a previous study [10] that the forecasting models are "not well suited for long-term predictions." However, we have introduced a forecasting model that *can* provide *highly accurate long-term predictions* both for the number of deaths caused by COVID-19 and the number of reported infected cases. Our success was based on replacing the usual logistic formula used in epidemiology by slightly more general formulas, namely, what we have called rational and birational formulas. Interestingly, these formulas are, in a sense, optimal, since elaborate "deep learning" algorithms could *not* improve the predictions obtained by using these formulas [11].

By training a formula to use data for the accumulative number of deaths in a given country until May 1, 2020, we were able to make predictions that were valid well beyond the end of the first wave, namely, for a period of more than 3.5 months. As shown in Figure 1, in Italy, there was no deviation between the curve depicting the number of deaths and the curve of our predictions, whereas in Germany, a small deviation began to occur in the second week of August 2020. In Spain, a larger deviation began to occur in the first week of August 2020 [12]. Our predictions for the number of reported infected cases were equally accurate [11].

These graphs show that following the easing of the first lockdown, the number of deaths did not increase in several European countries for approximately two months. Figure 1 presents three graphs depicting the total number of deaths caused by the COVID-19 epidemics in Italy, Spain, and Germany. These graphs compare the red curves indicating *real data* from the period of May 2 to November 2, 2020, with the green curves indicating the *predictions* made with an explicit mathematical formula introduced previously [11]. This formula contains 4 explicit constants that were determined by using data up to May 1, 2020. The green curves depicting this formula are indistinguishable from the red curves depicting the real data until mid-August 2020, despite the fact that the lockdown was lifted in the above countries by mid-June 2020.

XSL•FO
RenderX

Figure 1. Actual versus predicted cumulative number of reported deaths due to COVID-19 for Italy, Spain, and Germany. The constants of our mathematical model were determined using data up to May 1, 2020, which are depicted in light blue. The red curves correspond to the actual data up to November 2, 2020, which are the data during the period of our predictions. Our predictions are depicted in green.

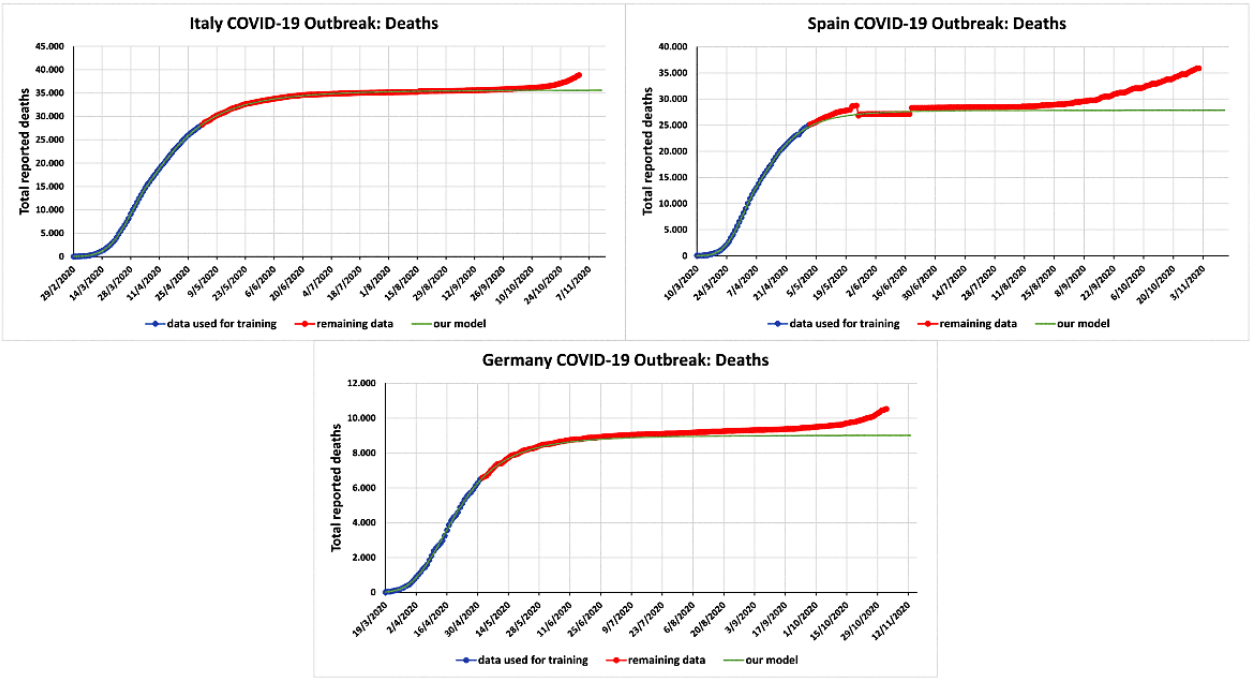
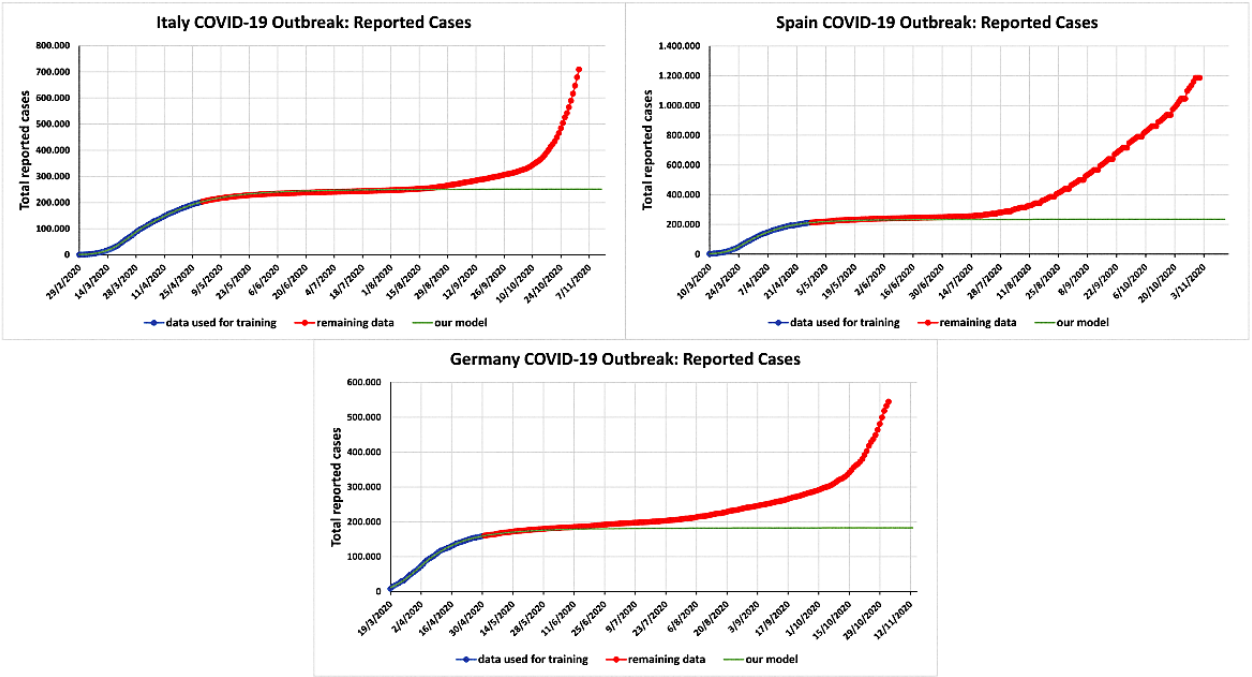


Figure 2 presents analogous curves for the number of reported infected cases. Comparing the red curves depicting the real data with the green curves depicting our predictions, it follows that until August-end for Italy, mid-July for Spain, and June-end for Germany, in 2020, the effect of the elimination of several restrictive measures had only a slight effect on the numbers of reported infected cases (compared with the numbers that would have occurred during the lockdown).

Figure 2. Actual versus predicted cumulative numbers of reported infected individuals with COVID-19 in Italy, Spain, and Germany. The prediction fits were obtained using data up to April 29, 2020, for Italy; April 27, 2020, for Spain; and April 30, 2020, for Germany. The relevant constants were determined with the data shown in light blue. The red curves correspond to the actual data up to November 2, 2020, which are the data during the period of our predictions. Our predictions are depicted in green.



XSL•FO
RenderX

Estimates for the Number of Reported Infected Cases and Deaths in Israel

We have applied our formulas to the cumulative number of deaths and reported infected cases for the third wave of COVID-19 in Israel. For the number of deaths, the rational formula predicted a plateau on April 8, 2021, with 3234 deaths and a total of 6060 deaths overall. The plateau is defined *as the point at whichthe rate of deaths for the given wave is 5% of the maximum rate.* For the number of reported infected cases in Israel, the rational formula predicted a plateau on April 16, 2021, with 497,992 cases for the third wave and a total of 825,399 cases overall. In Israel, the rate of infections of the third wave reached a maximum on January 20, 2021.

Higher Infectivity and Less Virulence

The predictions reported previously [13] were based on the assumption that *the characteristics of the virus remained unchanged.* The initial period following the easing of lockdown measures, when the number of deaths in the abovementioned countries did *not* increase and the number of reported infected cases was small, provided hope that the virulence and infectivity of the mutated form of the virus was much lower. However, the subsequent increase in the number of confirmed cases and deaths suggests an alternative explanation: the results reported previously [13] showed that if the easing of the restrictive measures leads up to doubling of the contacts of asymptomatic persons, the situation still remains essentially the same as it was during the lockdown period. Moreover, apparently, immediately after the lockdown period, the population at large continued to observe social distancing and other protective measures. Unfortunately, following this period of high vigilance, the population, especially younger individuals, thought that the danger was over and began enjoying the summer without observing the necessary protective measures.

A possible explanation for the observation that the number of deaths in the second wave was lower than the number expected given the very high number of infected individuals includes a number of factors. First, better medical practices and treatment are available. Second, older individuals and those with comorbidities are careful and avoid social contact. Indeed, the SARS-CoV-2 infection in younger individuals rarely leads to death. By extending the algorithm developed [13] into two subpopulations consisting of younger (below 40 years) and older (above 40 years) individuals, a previous study [9] on the epidemic in Greece found that if the number of contacts between asymptomatic “younger” persons increases, the number of reported infected cases increases but the number of deaths does not. In contrast, an increase in the number of contacts involving “older” persons leads to a dramatic increase in the number of deaths. Third and most importantly, it is possible that the mutated virus D614G was less virulent.

By directly computing the rate of change in the number of deaths in the second wave by using real data, it becomes clear that this rate is lower than the rate in the first wave, thereby supporting the suggestion that the mutated virus is less virulent. However, analogous computations for the rate of the change of the number of reported infected shows that this rate is much higher for the second wave, suggesting a much higher infectivity.

Concluding, it is worth noting that the period immediately following the easing of lockdown measures provides a very useful lesson. By using the well-publicized measures of social distancing, environmental hygiene, hand washing, and appropriate use of masks, it is possible to control the COVID-19 epidemic without imposing a strict lockdown.

Acknowledgments

ASF acknowledges support from EPSRC, UK (RG79707), in the form of a Senior Fellowship. The authors are grateful to their co-authors of the papers referred here, as well as to Dr E Gkrania-Klotsas for useful discussions. The authors acknowledge the contribution of the Open Access Team of the University of Cambridge for covering the Article Processing Fee.

Conflicts of Interest

None declared.

References

1. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev 2020;7(6):1012-1023 [FREE Full text] [doi: 10.1093/nsr/nwaa036]
2. MacLean OA, Orton RJ, Singer JB, Robertson DL. No evidence for distinct types in the evolution of SARS-CoV-2. Virus Evol 2020 Jan;6(1):veaa034 [FREE Full text] [doi: 10.1093/ve/veaa034] [Medline: 32817804]
3. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Sheffield COVID-19 Genomics Group, et al. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. Cell 2020 Aug 20;182(4):812-827.e19 [FREE Full text] [doi: 10.1016/j.cell.2020.06.043] [Medline: 32697968]
4. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat Commun 2020 Nov 25;11(1):5986 [FREE Full text] [doi: 10.1038/s41467-020-19818-2] [Medline: 33239633]

XSL•FO
RenderX

5. Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, et al. Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. mBio 2020 Oct 30;11(6):e02707-e02720 [FREE Full text] [doi: 10.1128/mBio.02707-20] [Medline: 33127862]
6. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, COG-UK Consortium, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 2021 Jan 07;184(1):64-75.e11 [FREE Full text] [doi: 10.1016/j.cell.2020.11.020] [Medline: 33275900]
7. Callaway E. The coronavirus is mutating - does it matter? Nature 2020 Sep;585(7824):174-177. [doi: 10.1038/d41586-020-02544-6] [Medline: 32901123]
8. Holmdahl I, Buckee C. Wrong but useful - What Covid-19 epidemiologic models can and cannot tell us. N Engl J Med 2020 Jul 23;383(4):303-305. [doi: 10.1056/NEJMp2016822] [Medline: 32412711]
9. Fokas AS, Cuevas-Maraver J, Kevrekidis P. Easing COVID-19 lockdown measures while protecting the older restricts the deaths to the level of the full lockdown. Sci Rep 2021 Mar 12;11(1):5839 [FREE Full text] [doi: 10.1038/s41598-021-82932-8] [Medline: 33712637]
10. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
11. Fokas AS, Dikaios N, Kastis GA. Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. J R Soc Interface 2020 Aug;17(169):20200494 [FREE Full text] [doi: 10.1098/rsif.2020.0494] [Medline: 32752997]
12. Fokas AS, Dikaios N, Kastis GA. Covid-19: predictive mathematical formulae for the number of deaths during lockdown and possible scenarios for the post-lockdown period. Proc R Soc A (forthcoming) 2021;477:20200745. [doi: 10.1098/rspa.2020.0745]
13. Fokas AS, Cuevas-Maraver J, Kevrekidis PG. A quantitative framework for exploring exit strategies from the COVID-19 lockdown. Chaos Solitons Fractals 2020 Nov;140:110244 [FREE Full text] [doi: 10.1016/j.chaos.2020.110244] [Medline: 32863608]

Abbreviations

SEIR: susceptible-exposed-infected-recovered

Edited by C Basch; submitted 11.07.20; peer-reviewed by G Neofotistos, S Gohari; comments to author 28.10.20; revised version received 05.11.20; accepted 24.03.21; published 18.05.21.

Please cite as:
Fokas AS, Kastis GA
SARS-CoV-2: The Second Wave in Europe
J Med Internet Res 2021;23(5):e22431
URL: https://www.jmir.org/2021/5/e22431
doi:10.2196/22431
PMID:33939621

©Athanassios S Fokas, George A Kastis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.05.2021. 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

Experiences and Attitudes of Elementary School Students and Their Parents Toward Online Learning in China During the COVID-19 Pandemic: Questionnaire Study

Shu Cui[1,2], MD; Chao Zhang[3], MD; Shijiang Wang[1], MD; Xingong Zhang[4], MD; Lei Wang[1,2], MD; Ling Zhang[1,2], MD; Qiuyu Yuan[1,2], MD; Cui Huang[1,2], MD; Fangshuo Cheng[5], MD; Kai Zhang[1,2*], MD, PhD; Xiaoqin Zhou[1,2*], MD

[1]Department of Psychiatry, Chaohu Hospital, Anhui Medical University, Chaohu, Hefei, China

[2]School of Mental Health and Psychological Sciences, Anhui Medical University, Hefei, China

[3]Department of Pediatrics, Fuyang People's Hospital, Fuyang, China

[4]Department of Psychiatry, The Third People's Hospital of Fuyang, Fuyang, China

[5]Department of Psychiatry, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, China

[*]these authors contributed equally

Corresponding Author:
Xiaoqin Zhou, MD
Department of Psychiatry
Chaohu Hospital
Anhui Medical University
64 Chaohu Road
Chaohu, Hefei
China
Phone: 86 055182324014
Email: zhouxqlulu@126.com

Abstract

Background: Due to widespread SARS-CoV-2 infection, an emergency homeschooling plan was rigorously implemented throughout China.

Objective: This study aimed to investigate the experiences and attitudes of elementary school students and their parents (two generations from the same family) toward online learning in China during the pandemic.

Methods: A 16-item questionnaire was distributed at the 10-day and 40-day marks after the start of the first online course to 867 parent-child pairs and 141 parent-child pairs, respectively. The questionnaire was comprised of questions pertaining to course and homework completeness, effectiveness, reliability, and abundance as well as the students' enthusiasm for taking part in online classes and their satisfaction with the courses.

Results: Our findings indicate that 90.7% (786/867) of students exhibited high or moderate enthusiasm for participating in online classes. However, most students performed poorly in online learning classes and after-school homework. With regard to satisfaction, parents' and students' average scores were 7.35 and 7.25, respectively (10-point scoring system). During the second stage of this study, parents' positive evaluations for online learning declined, including those for the effectiveness and reliability of the courses. Furthermore, the proportion of students who completed the courses and homework on time decreased; this difference proved statistically significant (P=.047). The parents' and students' overall satisfaction with online learning also declined during the second stage (parents: 7.21; students: 7.23); however, the difference in overall satisfaction between the two stages was not statistically significant (parents: P=.53; students: P=.60). Several of the parents (315/867, 36.2%) indicated that assisting with and supervising the students' online learning resulted in increased stress. Further, 36% of parents expressed dissatisfaction with or provided suggestions for online learning; most parents and students hoped to return to face-to-face classes (parents: 823/867, 94.9%; students: 811/867, 93.5%). Finally, our results presented the following six main issues that parents were the most concerned about: (1) disappointment regarding timely interaction in courses; (2) apprehensiveness about students' understanding of the course; (3) the increased burden of annoying adult responsibilities; (4) concern about children's eyesight; (5) the idea that teachers' explanations were not detailed enough; and (6) concerns about the decline of students' interest in and attention toward online courses.

XSL•FO
RenderX

Conclusions: Online learning can prevent the spread of infectious diseases while still allowing elementary school students to attain knowledge. However, in our study, children's completion of the courses and homework were not satisfactory. Furthermore, their parents often experienced stress and had many concerns and complaints. Measures such as increasing the interactivity of the courses and prohibiting teachers from assigning tasks to parents could improve the effectiveness of these courses and the mental health of parents and students.

(J Med Internet Res 2021;23(5):e24496) doi:10.2196/24496

KEYWORDS

attitude; elementary school students; parents; online learning; COVID-19

Introduction

Due to widespread SARS-CoV-2 infection, the Chinese government postponed the opening of schools after the Spring Festival to prevent further infections [1]. Face-to-face socializing was also prohibited. China's Ministry of Education estimated that more than 270 million students were confined to their homes, including 17.67 million elementary school students [2].

The Ministry of Education stipulated that even though schools were closed, teaching must continue during the lockdown period [3]. Accordingly, online teaching has been rigorously implemented in China [4]. Since mid-February 2020, schools and teachers of all levels have made considerable efforts toward creating and delivering online courses via internet-based methods or television broadcasts [5]. Consequently, this has resulted in the largest online learning campaign in human history.

Previous research has shown that online education has great potential for addressing the availability and efficiency of education [6,7]. However, by itself, online education is not more effective than a classroom-based approach, and its effectiveness depends on how well instructional designs are integrated into effective learning principles [8]. Differences in content quality, interactivity, and platform availability may affect learning satisfaction [9,10], but it is not clear which aspects are the most important for online education in primary schools in China. Previous studies on problems related to online learning have focused primarily on college students [11-14]. However, few studies have focused on elementary school students' experiences and satisfaction with online learning. Another limitation that has been mentioned in previous studies is that researchers only assess online learning satisfaction from students' perspectives. However, parents' opinions also influence students' satisfaction with learning and can inspire students to learn [10]. This is worrisome, as several factors (concentration, self-discipline, and related factors) can result in a host of problems during online education [15,16].

This study aimed to investigate the experiences and attitudes of Chinese elementary school students and their parents toward online learning during the COVID-19 pandemic. Furthermore, this study evaluated differences in parents' satisfaction with online education between the 10-day and 40-day marks after the start of the first online course.

Methods

Ethical Approval

The study protocol was approved by the Research Ethics Committee of Chaohu Hospital of Anhui Medical University (approval number: 202001-kyxm-07). The enrolled participants received detailed explanations about this study and signed electronic informed consent forms (parents' consent for students' participation). All participants' personal information was kept confidential, including their names and internet protocol addresses.

Survey Questionnaire

The survey questionnaire was designed to assess the online learning experiences and attitudes of Chinese elementary school students and their parents (Multimedia Appendix 1) based on the concerns of parents of elementary school students and previous questionnaires [17]. Each questionnaire was completed by 1 student and 1 parent. In order to prevent selection bias from affecting the study outcomes, each parent was selected at random and chosen by their family. The questionnaire consisted of 16 items and focused primarily on obtaining basic information, including children's grades and the equipment that was used during online classes. Thereafter, a broader selection of information was gathered, including participants' levels of enthusiasm for online learning, the completion of online classes, the completion of assigned homework, the pressure on parents, and related factors. Items 1-12 were answered by parents and items 13-15 were answered by elementary school students. Item 16 was an open comment that was directed at parents; it was designed to obtain their opinions on online learning. To measure satisfaction, we used a 10-point scoring system (ranged from 1 to 10 with intervals of 1); 1 represented the lowest degree of satisfaction, and 10 represented the highest degree of satisfaction. The questionnaire was written in Chinese and was not translated into any other language. Primary education in China is compulsory for all children who reach a certain age (6-7 years). Primary school students are usually between the ages of 6 and 13 and are in grades 1-6. With regard to our questionnaire, primary school students only had to answer three simple questions, which they understood and correctly answered.

We pretested the questionnaire with 5 parent-child pairs of primary school students (not part of the research team) and 7 psychologists. They pretested the questionnaire to determine the feasibility and understanding of the questions and words and to provide feedback. The content validity of the final version of the questionnaire was 0.86. The Cronbach α of the

XSL•FO
RenderX

questionnaire was .73, which was within the appropriate, acceptable Cronbach α range (.70-.95) [18,19]. On the basis of the Kendall sample size calculation method [20], the minimum sample size had to be 10 times the number of items in the questionnaire plus 20% of the number of invalid questionnaires. Therefore, since the scale was composed of 16 items, the minimum sample size of this study had to be 192. Our study obtained 1008 valid questionnaires and therefore met the sample size requirements.

Survey Protocol

The questionnaire was produced and distributed by the authors. The relevant data were subsequently collected with the web-based survey tool Questionnaire Star (Ranxing Information Technology Company, Limited), a professional, web-based survey evaluation platform [21]. Questionnaire Star can be used to design questionnaires, collect data, create custom reports, and analyze results. We sent a questionnaire link to potential participants via WeChat (TenCent Holdings Limited), which is the most widely used social media platform in China.

Eligible participants included any Chinese elementary school students who participated in online education during the COVID-19 pandemic and their parents. The questionnaire survey was conducted during two separate phases in this study. A 16-item questionnaire was distributed at the 10-day and 40-day mark after the first online course. In the first phase, the questionnaire was sent to 867 parent-child pairs (867 elementary students and their parents). In the second phase, the questionnaire was sent to 141 parent-child pairs.

Statistical Analysis

Participants' responses were proportionally expressed and recorded with a Likert scale that was divided into "good," "average," and "poor" responses or "yes" and "no" responses. Continuous variables (ie, satisfaction scores) were compared with the Student *t* test. Categorical variables were compared with either chi-square tests or Fisher exact tests. Data were analyzed using PASW (Predictive Analytics SoftWare) Statistics 20 (IBM Corporation). *P* values of <.05 were considered statistically significant.

Data Availability

All the data that support our findings are presented in the manuscript. The data sets used and analyzed during this study can be made available from the corresponding author upon reasonable request.

Results

Factors Affecting Students' and Parents' Perceptions of Online Learning

The total number of participants included 867 parent-child pairs (1 parent for each child)—867 elementary school children and 867 respective parents—during the first stage of this study. During the second stage (30 days after the first interview), only 141 parent-child pairs (1 parent for each child) were included in this study. The majority of students (304/867, 35.1%) were grade 4 students (Table 1).

With regard to the equipment used in online classes, lower-grade students were more likely to choose a television. However, grade 6 students' following equipment choices exhibited relatively similar ratios: mobile phone (8/31, 25.8%), tablet (7/31, 22.6%), PC (8/31, 25.8%), and television (8/31, 25.8%; Table 1).

Table 1 shows the enthusiasm of primary school students in grades 1-6 who participated in online learning courses. The results revealed that most students (24/31, 37.4%) had enthusiasm for engaging in online learning courses. Surprisingly, 22.6% (7/31) of grade 6 students were not enthusiastic about taking online learning courses. This was statistically significant when compared to the enthusiasm of students in other grades (*P*=.006; Table 1).

With regard to completeness, two subitems were developed and pertained to online learning courses and the accompanying homework. Surprisingly, many of the students did not do well in the online classes or the after-class homework (Table 1). Notably, the degree to which online classes were completed was higher in both grade 1 (23/46, 50%) and grade 6 (13/31, 41.9%) than those in other grades; the difference proved statistically significant (*P*=.047; Table 1). Grade 6 students completed the largest amount of homework, while nearly half of the overall students performed well (15/31, 48.4%) (Table 1).

In this study, we designed the following three subcategories of evaluation through which the parents of elementary school students could evaluate the quality of online courses: effectiveness, reliability, and abundance. The results show that the majority of elementary school students' parents indicated that the reliability (649/867, 74.9%), effectiveness (334/867, 38.5%), and abundance (564/867, 65.1%) of online courses were good. However, the parents' views were inconsistent among the different grades. As such, more than 10% of grades 1 and 6 students rated the effectiveness of online classes as poor (5/46, 10.9%; 5/31, 16.1%). Furthermore, 16.1% (5/31) of the parents of grade 6 elementary students believed that the abundance of online courses was insufficient.

XSL•FO
RenderX

Table 1. Factors affecting student and parental perceptions of online learning at the first stage (10-day mark).

Factor	Grade 1 (n=46)	Grade 2 (n=207)	Grade 3 (n=129)	Grade 4 (n=304)	Grade 5 (n=150)	Grade 6 (n=31)	Chi-square (*df*)	*P* value
Equipment, n (%)							58.44 (5)	<.001
Mobile phone	7 (15.2)	82 (39.6)	33 (25.6)	109 (35.9)	50 (33.3)	8 (25.8)		
Tablet	16 (34.8)	14 (6.8)	20 (15.5)	33 (10.9)	20 (13.3)	7 (22.6)		
PC	6 (13)	14 (6.8)	8 (6.2)	20 (6.6)	13 (8.7)	8 (25.8)		
Television	17 (37)	97 (46.9)	68 (52.7)	142 (46.7)	67 (44.7)	8 (25.8)		
Enthusiasm, n (%)							24.494 (5)	.006
High	27 (58.7)	153 (73.9)	103 (79.8)	247 (81.2)	123 (82)	20 (64.5)		
Moderate	13 (28.3)	31 (15)	16 (12.4)	31 (10.2)	18 (12)	4 (12.9)		
Low	6 (13)	23 (11.1)	10 (7.8)	26 (8.6)	9 (6)	7 (22.6)		
Course completeness, n (%)							18.536 (5)	.047
Good	8 (17.4)	18 (8.7)	11 (8.5)	26 (8.6)	10 (6.7)	5 (16.1)		
Average	15 (32.6)	39 (18.8)	23 (17.8)	47 (15.5)	28 (18.7)	8 (25.8)		
Poor	23 (50)	150 (72.5)	95 (73.6)	231 (76)	112 (74.7)	18 (58.1)		
Homework completeness, n (%)							17.113 (5)	.07
Good	5 (10.9)	17 (8.2)	14 (10.9)	21 (6.9)	9 (6)	7 (22.6)		
Average	7 (15.2)	36 (17.4)	19 (14.7)	51 (16.8)	34 (22.7)	8 (25.8)		
Poor	34 (73.9)	154 (74.4)	96 (74.4)	232 (76.3)	107 (71.3)	16 (51.6)		
Course effectiveness, n (%)							10.381 (5)	.41
Good	19 (41.3)	79 (38.2)	53 (41.1)	116 (38.2)	59 (39.3)	8 (25.8)		
Average	21 (45.7)	129 (58)	66 (51.1)	166 (54.6)	81 (54)	19 (61.3)		
Poor	6 (13)	8 (3.9)	10 (7.8)	22 (7.2)	10 (6.7)	4 (12.9)		
Course reliability, n (%)							23.046 (5)	.01
Good	34 (73.9)	153 (73.9)	98 (76)	236 (77.6)	110 (73.3)	18 (58.1)		
Average	8 (17.4)	53 (25.6)	26 (20.2)	58 (19.1)	35 (23.3)	9 (29)		
Poor	4 (8.7)	1 (0.5)	5 (3.9)	10 (3.3)	5 (3.3)	4 (12.9)		
Course abundance, n (%)							25.599 (5)	.004
Good	25 (54.3)	136 (65.7)	91 (70.5)	207 (68.1)	94 (62.7)	11 (35.5)		
Average	17 (37)	61 (29.5)	30 (23.3)	87 (28.6)	51 (34)	15 (48.4)		
Poor	4 (8.7)	10 (4.8)	8 (6.2)	10 (3.3)	5 (3.3)	5 (16.1)		

Parents' Perceived Pressure From and Satisfaction With Online Learning

This study assessed the pressures that parents had to deal with during their children's online education. This study also measured parents' satisfaction with online learning during the COVID-19 outbreak (Table 2).

As indicated in Table 2, the parents of lower-grade students were under higher levels of pressure than parents of higher-grade students. The parents of grade 1 students (high pressure: 21/46, 45.7%) were generally the most stressed about their children's online lessons (Table 2).

With regard to satisfaction, most of the parents (675/867, 77.9%) were satisfied with the online learning courses; they scored above 6 points on the satisfaction scale (10-point scoring system; Table 2). In accordance with their parents, most of the students (641/867, 73.9%) were satisfied with their online learning courses; they also scored above 6 points on the satisfaction scale. Grade 6 students and their parents were the least satisfied with online learning, followed by grade 1 students and their parents. Interestingly, although the difference was not statistically significant (*P*=.053), grade 6 students reported higher satisfaction scores than their parents.

The results indicated that most of the parents (823/867, 94.9%) and students (811/867, 93.5%) hoped to return to face-to-face learning in their future studies (Table 2). Interestingly, 16.1% (5/31) of grade 6 students wanted to continue attending online classes in the future.

XSL•FO
RenderX

Table 2. Parents' perceived pressure from and satisfaction with online learning during the first stage (10-day mark) of this study.

Variable	Grade 1 (n=46)	Grade 2 (n=207)	Grade 3 (n=129)	Grade 4 (n=304)	Grade 5 (n=150)	Grade 6 (n=31)	Chi-square (*df*)	*P* value
Parents' pressure, n (%)							23.902 (5)	.008
Low	9 (19.6)	87 (42)	56 (43.4)	124 (40.8)	82 (54.7)	12 (38.7)		
Average	16 (34.8)	48 (23.2)	22 (17.1)	65 (21.4)	22 (14.7)	9 (29)		
High	21 (45.7)	72 (34.8)	51 (39.5)	115 (37.8)	46 (30.7)	10 (32.3)		
Parents' future choice, n (%)							11.851 (5)	.04
Online learning	4 (8.7)	9 (4.3)	3 (2.3)	11 (3.6)	14 (9.3)	3 (9.7)		
School	42 (91.3)	198 (95.7%)	126 (97.7)	293 (96.4)	136 (90.7)	28 (90.3)		
Parents' satisfaction score, mean (SD)	6.83 (2.46)	7.64 (2.02)	7.12 (2.73)	7.33 (2.46)	7.51 (2.07)	6.55 (2.58)	2.198 (5)	.053
Students' preferred online learning, n (%)							3.430 (5)	.63
Yes	25 (54.3)	92 (44.7)	67 (51.9)	153 (50.3)	78 (52)	17 (54.8)		
No	21 (45.7)	114 (55.3)	62 (48.1)	151 (49.7)	72 (48)	14 (45.2)		
Students' future choice, n (%)							16.517 (5)	.006
Online learning	3 (6.5)	10 (4.8)	8 (6.2)	12 (3.9)	18 (12)	5 (16.1)		
School	43 (93.5)	197 (95.2)	121 (93.8)	292 (96.1)	132 (88)	26 (83.9)		
Students' satisfaction score, mean (SD)	6.78 (2.29)	7.47 (2.35)	7.16 (2.16)	7.23 (2.40)	7.30 (2.28)	6.61 (2.67)	1.170 (5)	.32

The Attitudes of Elementary School Students and Their Parents During the Follow-up

This study was divided into two stages. The first stage of the investigation commenced 10 days after the online course began, while the second stage started 40 days after the course began. There were no significant differences in the elementary students' and parents' equipment use (P=.35), enthusiasm (P=.73), stress (P=.96), or satisfaction (students: P=.60; parents: P=.53) between the two phases.

With regard to completeness, fewer students completed their courses (11/141, 7.8%) and after-class homework (11/141, 7.8%) during the second stage compared to those in the first stage (completed course: 238/867, 27.5%; completed homework: 228/867, 26.3%). This difference proved statistically significant (completed course: P<.001; completed homework: P<.001; Table 3). Furthermore, the parents indicated that the quality of the online courses in the second stage (effectiveness: 130/141, 92.2%; reliability: 133/141, 94.3%) was lower than that in the first stage (effectiveness: 807/867, 93.1%; reliability: 838/867, 96.7%). The difference in the number of participants who believed that courses were reliable proved statistically significant (P=.01; Table 3).

Parents' and students' satisfaction levels for the online courses decreased during the second stage (parents' satisfaction: mean 7.21, SD 2.41; students' satisfaction: mean 7.13, SD 2.45) when compared to those in the first stage's survey (parents' satisfaction: mean 7.35, SD 2.35; students' satisfaction: mean 7.25, SD 2.43); however, the difference between the two stages was not statistically significant (parents' satisfaction: P=.53; students' satisfaction: P=.60; Table 3).

XSL•FO
RenderX

Table 3. The attitudes of elementary school students and their parents during the follow-up.

Variable	Baseline (n=867)	Follow-up (n=141)	Chi-square (*df*)	*P* value
Equipment, n (%)			3.272 (1)	.35
Mobile	289 (33.3)	41 (29.1)		
Tablet	110 (12.7)	16 (11.3)		
PC	69 (8)	8 (5.7)		
Television	399 (46)	76 (53.9)		
Enthusiasm, n (%)			0.643 (1)	.73
High	673 (77.6)	112 (79.4)		
Moderate	113 (13)	15 (10.6)		
Low	81 (9.3)	14 (9.9)		
Completion, n (%)				
Course			26.130 (1)	<.001
Good	78 (9)	1 (0.7)		
Average	160 (18.5)	10 (7.1)		
Poor	629 (72.5)	130 (92.2)		
Homework			23.277 (1)	<.001
Good	73 (8.4)	2 (1.4)		
Average	155 (17.9)	9 (6.4)		
Poor	639 (73.7)	130 (92.2)		
Course quality, n (%)				
Effectiveness			.358 (1)	.84
Good	334 (38.5)	51 (36.2)		
Average	473 (54.6)	79 (56)		
Poor	60 (6.9)	11 (7.8)		
Reliability			8.715 (1)	.01
Good	649 (74.9)	89 (63.1)		
Average	189 (21.8)	44 (31.2)		
Poor	29 (3.3)	8 (5.7)		
Abundance			0.731 (1)	.69
Good	564 (65.1)	96 (68.1)		
Average	261 (30.1)	40 (28.4)		
Poor	42 (4.8)	5 (3.5)		
Parent factors				
Pressure, n (%)			0.083 (1)	.96
Low	370 (42.7)	62 (44.)		
Average	182 (21)	29 (20.6)		
High	315 (36.3)	50 (35.5)		
Future choice, n (%)			1.339 (1)	.25
Online learning	44 (5.1)	4 (2.8)		
School	823 (94.9)	137 (97.2)		
Satisfaction score, mean (SD)	7.35 (2.35)	7.21 (2.41)	0.632 (1)	.53
Student factors				
Preferred online learning, n (%)			0.955 (1)	.33

XSL•FO
RenderX

Variable	Baseline (n=867)	Follow-up (n=141)	Chi-square (*df*)	*P* value
Yes	432 (49.8)	64 (45.4)		
No	435 (50.2)	77 (54.6)		
Future choice, n (%)			0.126 (1)	.72
Online learning	56 (6.5)	8 (5.7)		
School	811 (93.5)	133 (94.3)		
Satisfaction score, mean (SD)	7.25 (2.43)	7.13 (2.45)	0.525 (1)	.60

Parents' Open Comments Concerning Elementary School Students' Online Education

In the open comments, participants (parents) indicated that online classes effectively used their time and network so that classes were not suspended during the COVID-19 pandemic. In terms of deficiency, parents mentioned the following six main issues: (1) disappointment regarding timely interaction in online courses; (2) worry about students not understanding the course; (3) the increased burden of annoying adult responsibilities; (4) concern regarding children's eyesight; (5) concern that teachers' explanations were not detailed enough; and (6) concern about the decline of students' interest and attention toward online courses. We summarize the details in Textbox 1.

Textbox 1. Summary of parents' open comments. In total, 73% (736/1008) of parents answered the open questions.

Top question

- In total, 18.7% (188/1008) of parents thought that the interactions during the classes were inadequate.
- These parents stated that because online educational videos were taped in advance, there was a lack of question-and-answer interactions between teachers and students.
- These parents suggested that measures should be taken to ensure that teachers are aware of children's questions so that they can respond to specific questions or correct children's mistakes.

Second highest ranked question

- In total, 15.2% (153/1008) of parents were concerned that children could not understand the content of online educational videos.

Third highest ranked question

- In total, 13.6% (137/1008) of parents complained that teachers' demands, including monitoring children's online studies, checking homework, and regularly providing feedback on students' learning, greatly increased their workload, stress, and annoyance.
- A few parents were poorly educated and could not check their children's homework.

Fourth highest ranked question

- In total, 12.4% (125/1008) of parents were worried that prolonged exposure to electronic screens would lead to reduced eyesight in their children.

Fifth highest ranked question

- In total, 12.1% (122/1008) of parents thought that the online class durations were too short and that the teachers' explanations were not detailed enough.
- Only 2 parents felt that the online class durations were too long.

Sixth highest ranked question

- In total, 3.7% (37/1008) of parents claimed that online teaching lacks a learning and competitive atmosphere and that student's initiative and enthusiasm were not high.

Discussion

Principal Findings

The COVID-19 pandemic has radically changed many aspects of our lives. Furthermore, social distancing and restrictive movement policies have markedly derailed traditional educational practices [22-24]. Consequently, there is a pressing need to innovate and implement alternative education and assessment strategies [25,26]. However, the COVID-19 pandemic has provided an opportunity for the greater implementation of digital learning in elementary education that requires students to stay at home [27]. The convenience and flexibility provided by online classes seem to contribute to these classes' proliferation and popularity [28].

Although previous studies have asserted that learners gain slightly less knowledge in online environments [29-31], our survey results (in the study's first phase) showed that 93.1%

XSL•FO
RenderX

(807/867) of parents believed that the online courses were effective and were able to convey knowledge. Conversely, a study from Ghana found that only 40 (18.7%) of their respondents agreed that they were able to learn effectively at home, while 174 (81.3%) respondents disagreed with that statement [32]. These differences may be related to the different preparation times, study content, and equipment in online courses among different countries. There is an abundance of content for online learning courses in China, as the educational content of online courses was prepared early after the onset of the pandemic. Furthermore, the courses were designed so that students could use a variety of devices to participate, including students from families that do not have internet connections; they can still access the courses through their televisions. These measures have considerably increased the effectiveness of online learning in China. However, it is worth noting that during the second survey stage, the proportion of respondents who thought that online courses were effective decreased. This may have been due to long online lessons, which make it difficult for children to concentrate, thereby reducing their productivity. The number of participants in the second stage only consisted of about one-quarter (141/867, 16.3%) of the participants in the first stage. The reason for this may have been that parents' enthusiasm for the web-based survey declined during the second stage.

Satisfaction is a vital factor for determining the quality of online learning [33-35], as it reflects students' pleasure and fulfillment with the different aspects of learning services [36]. This study indicated that most parents (675/867, 77.9%) were satisfied with the online learning courses; they scored above 6 points on the satisfaction scale. In accordance with their parents, most students (641/867, 73.9%) were satisfied with the online learning courses; they also scored above 6 points on the satisfaction scale. Parents' and students' satisfaction with the online courses decreased during the second stage; however, this did not prove to be statistically significant (parents' satisfaction: P=.53; students' satisfaction: P=.60).

Grade 6 students and their parents were found to be the least satisfied with online learning, followed by grade 1 students and their parents. These participants felt that the courses were ineffective and unreliable and that the content was not abundant. Therefore, for primary school students and parents, curriculum quality was closely related to satisfaction. It is important to consider that grade 6 learners are under pressure due to the junior high school entrance examinations. This is notable because online classes only teach basic knowledge and do not allow for the conduction of extracurricular classes to improve exam scores. Grade 1 students experience cognitive pressure because of their recent transition to primary school from their carefree kindergartens. Therefore, they are more likely to develop adjustment disorders, which result in poor evaluations of a curriculum's quality.

Other studies on satisfaction have indicated that there are certain factors that affect students' satisfaction with online learning environments, such as their interactions and self-regulation [37,38]. Parahoo et al [39] indicated that the interactions among students, teachers, and classmates are an important dimension of students' satisfaction with online learning. Kuo and colleagues [40] found that learner-instructor and learner-content interactions are significant positive predictors of students' satisfaction. Another study's findings support the idea that learner-instructor interactions contribute to students' satisfaction [41]. In our survey's open comment section, the most frequently reported issue was the lack of interaction during online learning. This may account for the drop in satisfaction during the second survey phase. Sun and Chen [42] noted that students' main difficulties during online learning were staying motivated, adhering to schedules, and studying regularly. Unlike the first stage of our study, only 7.8% (11/141) of students completed their courses and finished their homework on time after a 1-day online learning class. Due to the psychological characteristics of children, few elementary students are able to consistently complete their online lessons and maintain self-discipline [43].

An advantage of our study is that we were able to assess the causes of parental anxiety related to online lessons. The factor that most frequently hindered students' learning was a lack of interaction, as identified by parents' open comments. Online courses are taped in advance, so there was a lack of timely, two-way interactions between students and teachers. Our results indicated that students generally received information passively and lacked active communication during their online classes. Furthermore, students often did not understand certain questions. Consequently, students may lose interest in online classes over time. Another potential problem is that long online courses may result in students becoming addicted to their computers and televisions. Furthermore, prolonged exposure to computers, mobile phones, or televisions can cause vision loss in elementary school students [44].

Our results indicated that a lack of interactivity may be the most important factor affecting Chinese primary school students' satisfaction with online courses. The online lessons in this study were recorded in advance, and the videos were played to primary school students later, which resulted in the one-way flow of teaching information. This is worrisome because clear explanations and communication for clarifying questions are especially important for distance learners. In contrast, online education in high-income countries has exhibited some improvement and enhancement. They emphasized more on interactivity and student participation and considered this factor when planning online courses. For example, Hrastinski [45] provided the following theory in his research:

> *If we want to enhance online learning, it needs to enhance online learners' participation and interactive experience.*

Suppan and colleagues [9] used a highly interactive online learning module that was tailored to customers' timely feedback and prevented content skipping. Their results showed that their module could enhance medical students' asynchronous distance learning in terms of knowledge acquisition. Synchronous e-learning based on interactive live webcasting has also been verified to be effective and feasible [46]. These results are consistent with our conclusions.

The low homework completion rates and high pressure on parents found in our study suggest that online learning tasks may be beyond the capacity of students and parents and may

XSL•FO
RenderX

cause parent-child conflicts and emotional problems. Our data are consistent with those of another web-based survey in China, in which parents' Self-rating Anxiety Scale results showed that the degree of anxiety was higher than normal. Additionally, 17.6% of students were suspected of experiencing emotional problems during online homeschooling [47]. Another survey showed that 73.9% of primary and secondary school students' parents felt that their burden increased, and compared to these parents, the burden on parents of primary school students in grades 1-3 increased by a higher degree (79.3%) [48].

In this study, 13.6% (137/1008) of parents complained about having to supervise their children, check their homework, and frequently deliver feedback to the teachers. This considerably increased parents' workload, stress, and annoyance. Moreover, several parents were unable to help their children, as they were uneducated. According to a previous study, there has been an alarming increase in child abuse and domestic violence rates in Brazil during the pandemic. This may be related to families' financial constraints, increased parental burdens resulting from school closures, parental stress, and the difficulty of dealing with children's irritability during isolation [49]. In our survey's open comment section, a parent wrote that when he was supervising his child, the child was undisciplined in class and perfunctory in completing his homework. The parent became particularly irritable and violent and stated that he even beat the child. The reason for such conflicts may be that parents endlessly nag their children when they are supervising their children's studies and correcting homework. This often results in children feeling that their space for independence is greatly compressed, which gives rise to conflicts between parents and students [50].

In academic circles, it is generally believed that there are utilitarian education and teaching concepts in China. Teachers who believe that "practice makes perfect" require students to perform many exercises during and after class. Since online courses in primary schools are prerecorded and lack teacher-student interaction, teachers transfer their responsibility of correcting homework to parents, which increases conflict rates and stress. Our results thus offer a new strategy for solving parent-child conflicts and emotional problems during online homeschooling.

Limitations

There are several limitations to this study that need to be discussed. First, the sample size was not very large. In future studies, a larger sample size should be used to validate this paper's results. Second, this study did not compare elementary school students' tests scores from before and after online learning. Test scores can provide a more intuitive perspective on the effects of online learning. However, due to the regulations of the Ministry of Education, we were unable to obtain the scores of the elementary school students. Third, our scale does not provide demographic data, such as age, gender, or participants' household incomes. As such, it was impossible to compare the differences among participants' demographic data. This is problematic because elementary school students of different ages, genders, or income levels may have different experiences and attitudes toward online learning. Lastly, we did not investigate teachers' attitudes toward online learning. These issues need to be explored in future research.

Conclusion

To the best of our knowledge, this study is the first to evaluate experiences and attitudes toward online learning among participants of two generations in the same family during the COVID-19 pandemic. Online learning can prevent the spread of infectious diseases and allow elementary school students to gain knowledge. Most enrolled elementary school students (673/867, 77.6% at baseline; 112/141,79.4% at follow-up) were very enthusiastic about participating in online classes, and students and their parents were satisfied with these classes. Students were able to adequately complete all of their lessons and after-school homework assignments during the initial phase of online learning. However, as time progressed, the percentage of students who completed their lessons and homework on time decreased. At this later stage, students' and parents' satisfaction with online lessons decreased. However, some online learning tasks may be beyond the capabilities of elementary school students and parents and may cause emotional and behavioral problems. This study provides evidence for policy changes that aim to reduce the amount of pressure on parents and improve mental health levels, including those that prohibit teachers from assigning the task of checking homework to parents and increase the amount of interaction between teachers and students in online classes.

Acknowledgments

We thank all the participants and corresponding authors. This study was funded by grants from the National Natural Science Foundation of China (grant 81801341) and the Anhui Provincial Key R&D Programme (grant 202004j07020030).

Authors' Contributions

KZ conceived and designed this study. XZ and KZ provided administrative support. SC, CZ, SW, XZ, LZ, LW, QY, CH, and FC collected and assembled the data. CS and KZ analyzed and interpreted the data. All authors wrote the manuscript and approved of the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Survey on the online education status of primary school students and the satisfaction of parents and students.
[DOCX File , 19 KB - jmir_v23i5e24496_app1.docx]

References

1. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc Health 2020 May;4(5):397-404 [FREE Full text] [doi: 10.1016/S2352-4642(20)30095-X] [Medline: 32272089]
2. Overview of education in China. China Education Center. URL: https://www.chinaeducenter.com/en/cedu.php [accessed 2021-04-29]
3. Zhang W, Wang Y, Yang L, Wang C. Suspending classes without stopping learning: China's education emergency management policy in the COVID-19 outbreak. Journal of Risk and Financial Management 2020 Mar 13;13(3):55 [FREE Full text] [doi: 10.3390/jrfm13030055]
4. Wang G, Zhang Y, Zhao J, Zhang J, Jiang F. Mitigate the effects of home confinement on children during the COVID-19 outbreak. Lancet 2020 Mar 21;395(10228):945-947 [FREE Full text] [doi: 10.1016/S0140-6736(20)30547-X] [Medline: 32145186]
5. Huang R, Tlili A, Chang TW, Zhang X, Nascimbeni F, Burgos D. Disrupted classes, undisrupted learning during COVID-19 outbreak in China: application of open educational practices and resources. Smart Learning Environments 2020 Jul 6;7:1-15 [FREE Full text] [doi: 10.1186/s40561-020-00125-8]
6. Bientzle M, Hircin E, Kimmerle J, Knipfer C, Smeets R, Gaudin R, et al. Association of online learning behavior and learning outcomes for medical students: Large-scale usage data analysis. JMIR Med Educ 2019 Aug 21;5(2):e13529 [FREE Full text] [doi: 10.2196/13529] [Medline: 31436166]
7. Meinert E, Alturkistani A, Foley KA, Brindley D, Car J. Examining cost measurements in production and delivery of three case studies using e-learning for applied health sciences: Cross-case synthesis. J Med Internet Res 2019 Jun 04;21(6):e13574 [FREE Full text] [doi: 10.2196/13574] [Medline: 31165718]
8. Cook DA, Sorensen KJ, Hersh W, Berger RA, Wilkinson JM. Features of effective medical knowledge resources to support point of care learning: a focus group study. PLoS One 2013 Nov 25;8(11):e80318. [doi: 10.1371/journal.pone.0080318] [Medline: 24282535]
9. Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, et al. Asynchronous distance learning of the National Institutes of Health Stroke Scale during the COVID-19 pandemic (e-learning vs video): Randomized controlled trial. J Med Internet Res 2021 Jan 15;23(1):e23594 [FREE Full text] [doi: 10.2196/23594] [Medline: 33428581]
10. Chen T, Peng L, Yin X, Rong J, Yang J, Cong G. Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. Healthcare (Basel) 2020 Jul 07;8(3):200 [FREE Full text] [doi: 10.3390/healthcare8030200] [Medline: 32645911]
11. Peters MA, Wang H, Ogunniran MO, Huang Y, Green B, Chunga JO, et al. China's internationalized higher education during Covid-19: Collective student autoethnography. Postdigital Science and Education 2020 May 08;2:968-988 [FREE Full text] [doi: 10.1007/s42438-020-00128-1]
12. 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]
13. Crawford J, Butler-Henderson K, Rudolph J, Malkawi B, Glowatz M, Burton R, et al. COVID-19: 20 countries' higher education intra-period digital pedagogy responses. Journal of Applied Learning & Teaching 2020 Apr 1;3(1):9-28 [FREE Full text] [doi: 10.37074/jalt.2020.3.1.7]
14. Huang RH, Liu DJ, Tlili A, Yang JF, Wang HH. Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak. Beijing, China: Smart Learning Institute of Beijing Normal University; Mar 2020.
15. Cavanaugh CS, Barbour MK, Clark T. Research and practice in K-12 online learning: A review of open access literature. International Review of Research in Open and Distance Learning 2009 Feb;10(1):1-22 [FREE Full text] [doi: 10.19173/irrodl.v10i1.607]
16. Gerstein J. Beyond the game: Quest Atlantis as an online learning experience for gifted elementary students. J Virtual Worlds Res 2009 Mar 06;2(1):1-18. [doi: 10.4101/jvwr.v2i1.403]
17. Online class satisfaction survey. Star Q. URL: https://www.wjx.cn/jq/96759811.aspx [accessed 2021-04-29]
18. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
19. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997 Feb 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
20. Kendall M. Multivariate Analysis. London: Charles Griffin & Co. LTD; 1980.

XSL•FO
RenderX

21. Shi Y, Wang J, Yang Y, Wang Z, Wang G, Hashimoto K, et al. Knowledge and attitudes of medical staff in Chinese psychiatric hospitals regarding COVID-19. Brain Behav Immun Health 2020 Apr;4:100064 [FREE Full text] [doi: 10.1016/j.bbih.2020.100064] [Medline: 32289123]
22. Rose S. Medical student education in the time of COVID-19. JAMA 2020 Jun 02;323(21):2131-2132. [doi: 10.1001/jama.2020.5227] [Medline: 32232420]
23. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. Lancet Public Health 2020 May;5(5):e243-e244 [FREE Full text] [doi: 10.1016/S2468-2667(20)30084-0] [Medline: 32275858]
24. Chick RC, Clifton GT, Peace KM, Propper BW, Hale DF, Alseidi AA, et al. Using technology to maintain the education of residents during the COVID-19 pandemic. J Surg Educ 2020;77(4):729-732 [FREE Full text] [doi: 10.1016/j.jsurg.2020.03.018] [Medline: 32253133]
25. Bao W. COVID-19 and online teaching in higher education: A case study of Peking University. Hum Behav Emerg Technol 2020 Apr;2(2):113-115 [FREE Full text] [doi: 10.1002/hbe2.191] [Medline: 32510042]
26. Agarwal S, Kaushik JS. Student's perception of online learning during COVID Pandemic. Indian J Pediatr 2020 Jul;87(7):554 [FREE Full text] [doi: 10.1007/s12098-020-03327-7] [Medline: 32385779]
27. Basilaia G, Kvavadze D. Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. Pedagogical Research 2020;5(4):1-9 [FREE Full text] [doi: 10.29333/pr/7937]
28. Harasim L. Shift happens: online education as a new paradigm in learning. Internet High Educ 2000;3(1-2):41-61. [doi: 10.1016/s1096-7516(00)00032-4]
29. Platt CA, Raile ANW, Yu N. Virtually the same? Student perceptions of the equivalence of online classes to face-to-face classes. J Online Learn Teach 2014 Dec:489-503.
30. Neuhauser C. Learning style and effectiveness of online and face-to-face instruction. American Journal of Distance Education 2002;16(2):99-113. [doi: 10.1207/s15389286ajde1602_4]
31. Paechter M, Maier B. Online or face-to-face? Students' experiences and preferences in e-learning. Internet High Educ 2010 Dec;13(4):292-297. [doi: 10.1016/j.iheduc.2010.09.004]
32. Owusu-Fordjour C, Koomson CK, Hanson D. The impact of Covid-19 on learning-the perspective of the Ghanaian student. European Journal of Education Studies 2020;7(3):88-101. [doi: 10.5281/zenodo.3753586]
33. Shen D, Cho MH, Tsai C, Marra R. Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. Internet High Educ 2013 Oct;19:10-17. [doi: 10.1016/j.iheduc.2013.04.001]
34. Ke F, Kwak D. Online learning across ethnicity and age: A study on learning interaction participation, perception, and learning satisfaction. Comput Educ 2013 Feb;61:43-51. [doi: 10.1016/j.compedu.2012.09.003]
35. Palmer SR, Holt DM. Examining student satisfaction with wholly online learning. J Comput Assist Learn 2009 Apr;25(2):101-113. [doi: 10.1111/j.1365-2729.2008.00294.x]
36. Kauffman H. A review of predictive factors of student success in and satisfaction with online learning. Research in Learning Technology 2015;23:1-13 [FREE Full text] [doi: 10.3402/rlt.v23.26507]
37. Horzum MB. Interaction, structure, social presence, and satisfaction in online learning. EURASIA J MATH SCI T 2017 Jun 16;11(3):505-512 [FREE Full text] [doi: 10.12973/eurasia.2014.1324a]
38. Croxton RA. The role of interactivity in student satisfaction and persistence in online learning. J Online Learn Teach 2014 Jun;10(2):314-325 [FREE Full text]
39. Parahoo SK, Santally MI, Rajabalee Y, Harvey HL. Designing a predictive model of student satisfaction in online learning. Journal of Marketing for Higher Education 2015 Oct 21;26(1):1-19. [doi: 10.1080/08841241.2015.1083511]
40. Kuo YC, Walker AE, Belland BR, Schroder KEE, Kuo YT. A case study of integrating Interwise: Interaction, internet self-efficacy, and satisfaction in synchronous online learning environments. The International Review of Research in Open and Distance Leraning 2014;15(1):161-181 [FREE Full text] [doi: 10.19173/irrodl.v15i1.1664]
41. Alqurashi E. Predicting student satisfaction and perceived learning within online learning environments. Distance Education 2018 Dec 03;40(1):133-148. [doi: 10.1080/01587919.2018.1553562]
42. Sun A, Chen X. Online education and its effective practice: A research review. Journal of Information Technology Education: Research 2016;15:157-190 [FREE Full text] [doi: 10.28945/3502]
43. Kerr MS, Rynearson K, Kerr MC. Student characteristics for online learning success. Internet High Educ 2006;9(2):91-105. [doi: 10.1016/j.iheduc.2006.03.002]
44. Larreamendy-Joerns J, Leinhardt G. Going the distance with online education. Rev Educ Res 2006 Dec 01;76(4):567-605. [doi: 10.3102/00346543076004567]
45. Hrastinski S. A theory of online learning as online participation. Comput Educ 2009 Jan;52(1):78-82. [doi: 10.1016/j.compedu.2008.06.009]
46. Jones RB, Maramba I, Boulos MNK, Alexander T. Use of live interactive webcasting for an international postgraduate module in ehealth: case study evaluation. J Med Internet Res 2009 Nov 13;11(4):e46 [FREE Full text] [doi: 10.2196/jmir.1225] [Medline: 19914901]
47. Zhao Y, Guo Y, Xiao Y, Zhu R, Sun W, Huang W, et al. The effects of online homeschooling on children, parents, and teachers of Grades 1-9 during the COVID-19 pandemic. Med Sci Monit 2020 Sep 12;26:e925591 [FREE Full text] [doi: 10.12659/MSM.925591] [Medline: 32917849]

XSL•FO
RenderX

48. Du Y. Seventy percent of the parents of primary and secondary school students interviewed felt the burden of new courses started online. China Youth Daily. URL: http://news.youth.cn/sh/202006/t20200604_12354905.htm [accessed 2021-06-04]
49. Martins-Filho PR, Damascena NP, Lage RC, Sposato KB. Decrease in child abuse notifications during COVID-19 outbreak: A reason for worry or celebration? J Paediatr Child Health 2020 Dec;56(12):1980-1981 [FREE Full text] [doi: 10.1111/jpc.15213] [Medline: 33012011]
50. Toney LP, Kelley ML, Lanclos NF. Self- and parental monitoring of homework in adolescents: Comparative effects on parents' perceptions of homework behavior problems. Child Fam Behav Ther 2003;25(1):35-51. [doi: 10.1300/J019v25n01_03]

Abbreviations

PASW: Predictive Analytics SoftWare

Edited by C Basch; submitted 22.09.20; peer-reviewed by L Chang, N Kaur, A Khaleghi, Z Aghaei; comments to author 19.10.20; revised version received 09.12.20; accepted 16.04.21; published 19.05.21.

Please cite as:
Cui S, Zhang C, Wang S, Zhang X, Wang L, Zhang L, Yuan Q, Huang C, Cheng F, Zhang K, Zhou X
Experiences and Attitudes of Elementary School Students and Their Parents Toward Online Learning in China During the COVID-19 Pandemic: Questionnaire Study
J Med Internet Res 2021;23(5):e24496
URL: https://www.jmir.org/2021/5/e24496
doi:10.2196/24496
PMID:33878022

©Shu Cui, Chao Zhang, Shijiang Wang, Xingong Zhang, Lei Wang, Ling Zhang, Qiuyu Yuan, Cui Huang, Fangshuo Cheng, Kai Zhang, Xiaoqin Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Utilizing Health Behavior Change and Technology Acceptance Models to Predict the Adoption of COVID-19 Contact Tracing Apps: Cross-sectional Survey Study

Samuel Tomczyk[1], DrPhil; Simon Barth[1]; Silke Schmidt[1], DrPhil; Holger Muehlan[1], DrPhil

Department Health and Prevention, Institute of Psychology, University of Greifswald, Greifswald, Germany

Corresponding Author:
Samuel Tomczyk, DrPhil
Department Health and Prevention
Institute of Psychology
University of Greifswald
Robert-Blum-Straße 13
Greifswald,
Germany
Phone: 49 3834 420 3806
Fax: 49 3834 420 3812
Email: samuel.tomczyk@uni-greifswald.de

Abstract

Background: To combat the global COVID-19 pandemic, contact tracing apps have been discussed as digital health solutions to track infection chains and provide appropriate information. However, observational studies point to low acceptance in most countries, and few studies have yet examined theory-based predictors of app use in the general population to guide health communication efforts.

Objective: This study utilizes established health behavior change and technology acceptance models to predict adoption intentions and frequency of current app use.

Methods: We conducted a cross-sectional online survey between May and July 2020 in a German convenience sample (N=349; mean age 35.62 years; n=226, 65.3% female). To inspect the incremental validity of model constructs as well as additional variables (privacy concerns, personalization), hierarchical regression models were applied, controlling for covariates.

Results: The theory of planned behavior and the unified theory of acceptance and use of technology predicted adoption intentions (R^2=56%-63%) and frequency of current app use (R^2=33%-37%). A combined model only marginally increased the predictive value by about 5%, but lower privacy concerns and higher threat appraisals (ie, anticipatory anxiety) significantly predicted app use when included as additional variables. Moreover, the impact of perceived usefulness was positive for adoption intentions but negative for frequency of current app use.

Conclusions: This study identified several theory-based predictors of contact tracing app use. However, few constructs, such as social norms, have a consistent positive effect across models and outcomes. Further research is required to replicate these observations, and to examine the interconnectedness of these constructs and their impact throughout the pandemic. Nevertheless, the findings suggest that promulgating affirmative social norms and positive emotional effects of app use, as well as addressing health concerns, might be promising strategies to foster adoption intentions and app use in the general population.

(J Med Internet Res 2021;23(5):e25447) doi:10.2196/25447

KEYWORDS

mHealth; COVID-19; UTAUT1; UTAUT2; health behavior change; theory of planned behavior; contact tracing; app; model; technology acceptance; cross-sectional studies; social norms; health communication; privacy; anxiety

XSL•FO
RenderX

Introduction

Background

With the global spread of the COVID-19 pandemic, there have been numerous efforts to develop digital and mobile health (mHealth) solutions to combat the spread of infections [1,2], support quarantine and social isolation, and improve monitoring and communication surrounding the virus [3]. To this end, COVID-19 contact tracing apps were proposed as a way to (1) monitor and track infection chains, (2) provide immediate support and information in case of an infection or contact with an infected person, (3) and support persons in quarantine by monitoring health and tailoring information and preventive actions [1,3,4]. Overall, contact tracing via mobile apps aims to increase perceived safety and security of the population, and to contain infections, some apps also include additional educational information and news updates on governmental regulations [5].

However, these developments encounter several ethical and practical challenges (eg, [1,6,7]). From a technical point of view, for instance, flawless performance of an app is essential to avoid false positives and for it to be perceived as accurate, reliable, and trustworthy (ie, efficacious) [8]. Research shows that perceived performance efficacy of a digital technology predicts behavioral compliance when confronted with false-alarm–prone systems [9]; therefore, the rate of false-positive alarms should be as low as possible for tracing apps. Furthermore, the app has to be accessible across different areas and regions (eg, rural and urban areas) to guarantee successful preventive tracing [7]. In addition, ethical questions concerning the digital divide resurface in this context [6,7,10]. Since the digital divide characterizes differences in access and reach of digital technologies (ie, primary divide) as well as capabilities and habit of use (ie, secondary divide), it challenges tracing apps as public health measures [10]. Elderly people, for example, report less use of smartphones, and they feel less competent in smartphone use [11,12]. This is particularly challenging regarding contact tracing apps, as older people represent a risk group for COVID-19 infections with a higher chance of more severe trajectories [13]. Additionally, using tracing apps also implies an agreement to share health-related data (ie, infection status) via mobile apps with governmental institutions or other citizens, raising concerns surrounding data privacy and a potential breach as well as misuse of health information beyond COVID-19 purposes (cf "surveillance creep" [3,6,7]). Since the implementation of the European Union General Data Protection Regulation in 2018, privacy concerns have repeatedly been discussed as a key variable in influencing attitudes and adoption intentions [14], particularly in the context of mHealth apps.

Hence, lively debates among journalists, scientific experts, and policymakers about the importance of weighing privacy, individual data security, and societal public health needs in digitally tracing COVID-19 infections have led to a variety of diverse developments in mHealth (eg, see [5,15]). A review conducted in May 2020 [15] described 17 tracing apps in 15 different countries with varying degrees of data processing and protection—in September 2020, the Council of Europe already listed 52 contact tracing apps worldwide [16]. Yet only 3 of these 17 reviewed apps (COVIDSafe in Australia, The e-Rouska in the Czech Republic, and VirusRadar in Hungary) were protected by respective data protection laws and provided consistent information on data storage and use policies.

In general, tracing apps tend to follow either a centralized or decentralized data processing approach or a hybrid of both, where either governmental or service institutions receive, monitor, and administer app data via a central server structure (eg, in France), or communication relies on peer-to-peer technology such as Bluetooth (eg, in Germany), to exchange randomly generated codes between app users within a certain radius (eg, a proximity of a few meters) to trace contacts. Either way, once a person's infection is validated by a health agency, a warning can be sent to stored contacts (centralized approach) or recipients of codes (decentralized approach) to inform them of a potential infection. So far, advantages and disadvantages regarding systems architecture, perceived responsibility, security, and privacy have already been discussed, ranging from personal first-hand experience [17] to systematic reviews (eg, [18]). While further differences between these approaches are beyond the scope of this paper and are discussed in detail elsewhere (eg, systems architecture [19,20]), it is important to note that these processes may further affect adoption intentions and continued app use in the population, as they are assumed to be connected to personal attitudes, such as perceived control, and privacy concerns. In fact, Trang et al [21] found that a strong privacy design, as implemented in a decentralized approach, predicted app acceptance and adoption intentions among critics of the tracing app as well as undecided participants who represented a majority of the sample. Similarly, a cross-cultural study also reported privacy and security concerns as important barriers to tracing app use [22].

Use of Contact Tracing Apps in the General Population

Mindful of these technical and ethical challenges of tracing apps, a general question therefore is whether the public uses these apps as intended. Simulation studies report that at least 56% of a population needs to use an app for it to have a public health impact [4]. However, use rates appear to be much lower in the general population so far (eg, [19,20]), with the differences being discussed in terms of data processing approaches, societal technology acceptance, and institutional or legal commitment. The Corona Warn-App (which uses a decentralized approach) was launched as the official German coronavirus contact tracing app on June 16, 2020. Until May 7, 2021, the app was downloaded 27.5 million times, which accounts for about 33.1% of the general population [23,24]. Given this large discrepancy between needed and currently reported app use, behavioral health research can help to identify predictors of app use and derive recommendations for preventive practice to increase app use and address perceived barriers in the general population. To this extent, this study draws from the literature on health behavior change as well as (mobile) technology acceptance to explore tracing app use in the general population. While we are aware of several studies that inspect barriers or motives pertaining to tracing app use (eg, [21,22]), few empirical studies are theory based in that they utilize established health behavior or technology acceptance models

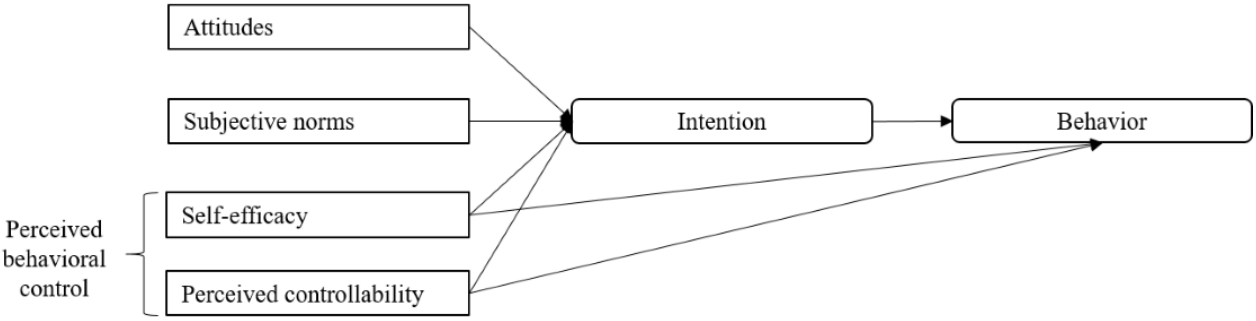
XSL•FO
RenderX

to investigate app use. So far, we have identified 3 studies based on the health belief model (HBM) [25], the protection motivation theory (PMT) [26], and the unified theory of acceptance and use of technology (UTAUT) [27]. While these studies investigated behavior change theories, they lack an assessment of social influence or social norms, even though these aspects are key factors of successful behavioral prevention during pandemics [28,29]. We believe that it is important to acknowledge the impact of social influence on contact tracing app use—as explicated in several health behavior change models.

Health Behavior Change

Ample research provides evidence of psychological processes connected to behavior change toward adaptive health behaviors in pandemics, such as keeping physical distance or practicing personal hygiene [28,30], which have been reiterated during the COVID-19 pandemic (eg, [30-32]). These processes are detailed in health behavior models like the HBM [33], Bandura's social cognitive theory [34], the PMT [35], and the theory of planned behavior (TPB) [36], which define several constructs associated with behavior change. According to the HBM, for instance, risk perception (eg, susceptibility to and fear of an illness) and perceived benefits (eg, reduced risk of infection) and barriers (eg, high costs, privacy concerns) influence behavior change tendencies. Social cognitive theory further adds the impact of efficacy expectancies [37] that comprise beliefs about the outcome of a behavior (ie, response efficacy, for instance, the belief that wearing masks significantly reduces infection risk) as well as one's ability to perform within a specific setting (ie, self-efficacy). The PMT combines these approaches by describing threat appraisals (risk perception) as well as coping appraisals (perceived benefits and barriers, response efficacy, and self-efficacy) as predictors of protection motivation and protective behavior. However, applied research points to the intention-behavior gap [38], describing a lack of implementation despite positive intentions, for instance, due to individual forgetfulness or a lack of opportunity to perform. Therefore, the TPB [36] further differentiates efficacy beliefs within the construct of perceived behavioral control (Figure 1).

Figure 1. Conceptual model of the theory of planned behavior (adapted from Ajzen [39]).



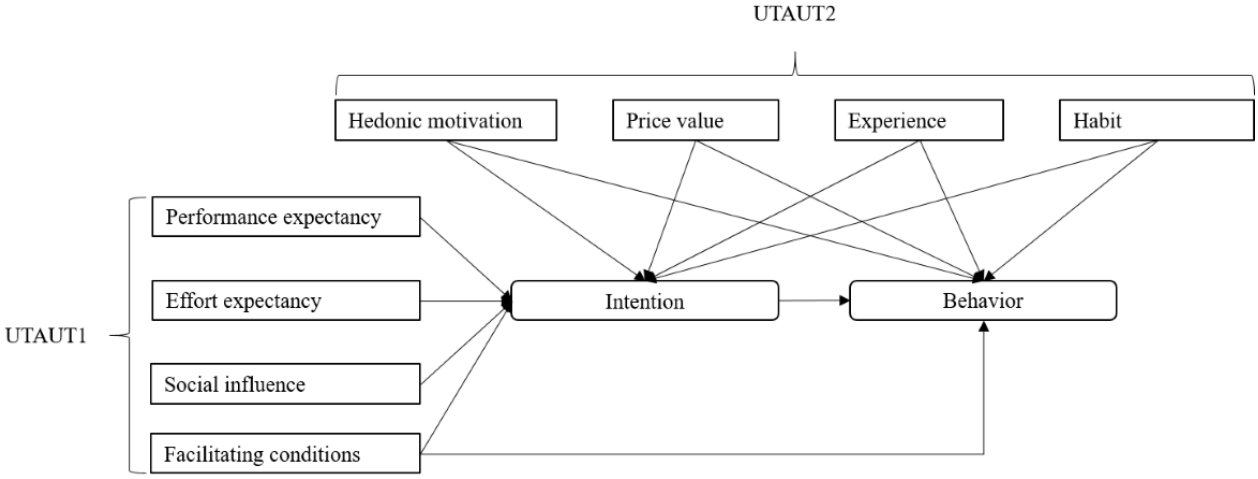
This construct comprises self-efficacy but also perceived controllability, which refers to the extent to which a person believes to be responsible for their behavioral performance within a specific setting [39]. For example, a person might report high self-efficacy regarding hand hygiene but low controllability due to a lack of soap or disinfectant available to them. Thus, the TPB assumes a direct link between perceived behavioral control and behavior as well as intentions. Other predictors of behavioral intention are attitudes and subjective norms. Attitudes are evaluative judgments of the target behavior, which might include risk perceptions and perceived benefits and barriers, and therefore connect with previously described health behavior models. Subjective norms address descriptive social influence (ie, how many people perform a target behavior) and injunctive social influence (ie, how many people suggest performing a target behavior) on individual intentions. Previous research has cemented the TPB as a popular and versatile framework for predicting health behaviors (eg, [40]) as well as use of mobile apps [41,42]. Current research on the COVID-19 pandemic underlines the predictive validity of the TPB for protective behaviors, with a particularly strong impact of subjective norms and efficacy beliefs [30,43,44]. Similarly, current research on the use of contact tracing apps also finds strong positive associations for response efficacy and moderate positive associations for self-efficacy [25-27]. Subjective norms were only measured in one study [27] and did not show significant associations with app use. However, the measure combined social norms, governmentally provided implementation support, and social support for app use as a composite measure of social influence, resulting in insufficient factorial and content validity.

Unified Theory of Acceptance and Use of Technology

While the TPB has received tremendous attention in health research, its popularity in mHealth research is rivaled by UTAUT [45]. The UTAUT itself is based on the technology acceptance model [46] and the theory of reasoned action [47], a precursor of the TPB, and it also describes psychosocial variables associated with the adoption of (new) technology [48-50]. These variables are performance expectancy, effort expectancy, social influence, and facilitating conditions in the first iteration (UTAUT1), with the addition of hedonic motivation, price value, habit, and experience in the second iteration (UTAUT2) [45,51] (Figure 2).

XSL•FO
RenderX

Figure 2. Conceptual model of the unified theory of acceptance and use of technology (UTAUT, first and second iterations) (adapted from Venkatesh et al [50]).



In contrast to the TPB, however, these descriptions are rather broad, and the model does not set strong theoretical or methodological boundaries for these variables, which has led to many publications but almost as many different iterations of the model, challenging the comparability of findings [50]. Nevertheless, the main implications of the UTAUT are that an adoption of a new technology is more likely if a person expects it to be useful, easy to access and use, recommended and supported by others, and if persons perceive themselves to be able to use it as intended. Moreover, if a person expects positive emotional reactions (eg, joy) and is familiar with similar technology use (ie, habit), this further increases the likelihood. Conversely, disadvantageous facilitating conditions, for example, information privacy concerns or low self-efficacy, can inhibit adoption intentions [25,52].

Conceptually, similarities to health behavior theories are apparent, as expectancies, social influences, and facilitating conditions (eg, self-efficacy, perceived barriers) play an important role in both approaches. Hence, multiple studies have used the UTAUT approach to predict adoption of mHealth apps in general [49], and more specifically, for instance, regarding mobile health care services among the elderly [53], or the use of COVID-19 contact tracing apps [27]. In their study on the acceptance of mHealth services, Sun et al [53] directly compared the predictive value of the TPB, the technology acceptance model (as a reduced version of the UTAUT), the PMT, and an integrated model, and found that an integrated model yields the best results, including positive effects (eg, social influence) and negative effects (eg, threat appraisals) at the same time. Across studies, performance expectancy has the strongest associations with adoption intentions, followed by effort expectancies and social influence. This is in line with current evidence on tracing app use but deviates from health behavior research that reports a similar impact of efficacy beliefs, but a stronger effect of social influence [30,43,44].

Moreover, in their comprehensive meta-analytic review of the UTAUT, Dwivedi et al [48] essentially confirm these observations, but they also point out that most UTAUT studies do not include user attitudes (ie, affective and cognitive evaluations of technology use). Instead, attitudinal assessments are often limited to expectancies. In terms of health behavior theories, this operationalization would exclude aspects of risk perception or threat appraisals (HBM, PMT), and attitudes (TPB), although attitudes were the strongest predictors of behavioral intentions in meta-analytic structural equation models of the UTAUT [48].

Furthermore, the role of social influence (ie, social norms) is not consistently defined in either the health behavior or the technology acceptance approach [54]. In fact, social norms can be categorized as descriptive and injunctive or prescriptive, and further defined as personal or societal, in that they refer to one's personal surroundings (eg, family or friends) or more general societal categories (eg, persons of the same age, the same country). In short, the reference frame defines the commitment and group orientation, which has been used by nudging interventions based on social identity models to foster technology use [55] or the uptake of vaccinations [56]. Using the general public as a reference can thus shape public behaviors (eg, wearing a mask in public), but does not necessarily affect personal beliefs as strongly (eg, private protective behaviors) [57]. In contrast, personal reference points can increase private behaviors more strongly. In the context of app use, it is therefore important to examine the differential impact of social norms.

Research Aims

In sum, this study aimed to compare the predictive value of the TPB, the UTAUT, an integrated model, and an extended model (cf [53]) regarding adoption intentions, and current use of COVID-19 infection tracing apps. By exploring both theories separately as well as concurrently, this study provides a blueprint for future research on contact tracing apps and other digital health technologies that combines health-related and technology-related perspectives.

First, we aimed to affirm previous mHealth studies on either the TPB or the UTAUT in the context of COVID-19 contact tracing apps and compared their tenability an integrative model.

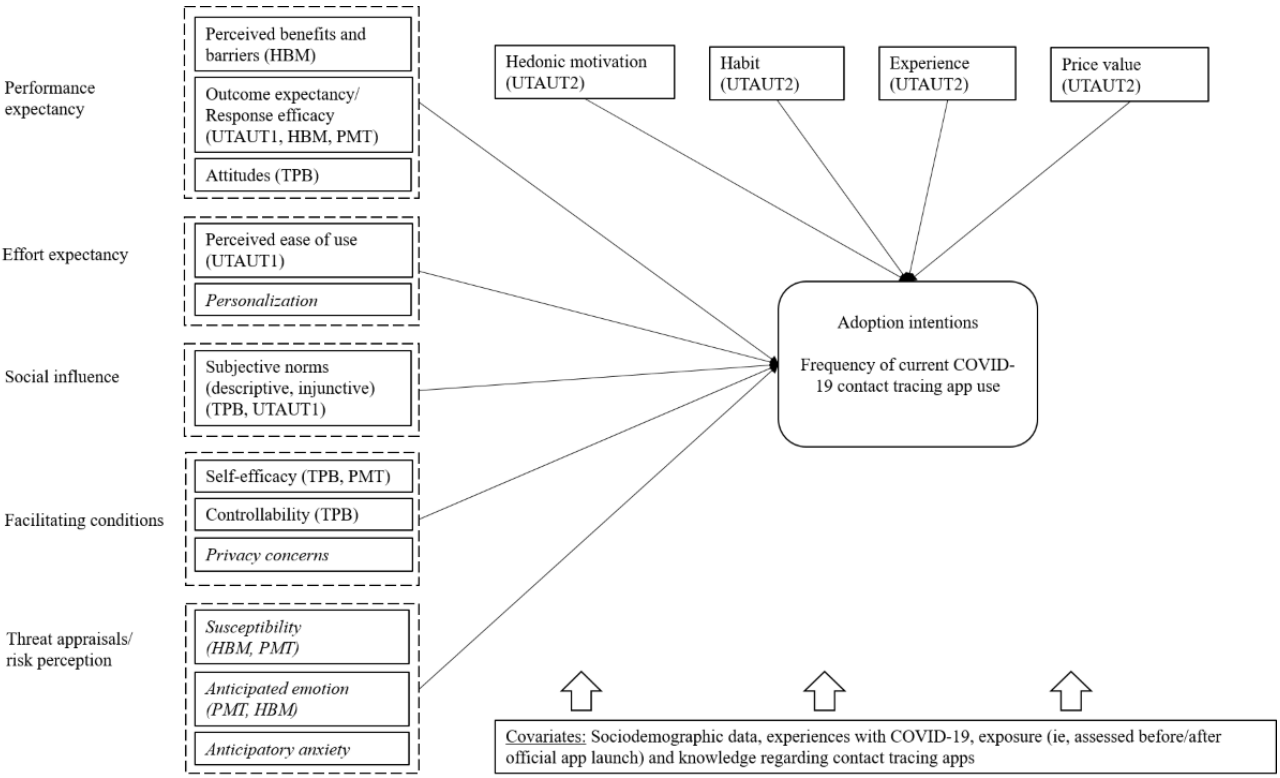
Second, while both models have been successfully used to predict health technology use, there are evident shortcomings to both approaches that can be overcome by combining them.

XSL•FO
RenderX

Therefore, in this study, we included barriers like privacy concerns, which are more common in technology-focused research (eg, [27,50]), into health behavior models, while also acknowledging the importance of threat appraisals, popularized in health behavior models [53].

Third, we aimed to expand upon research on either model by considering the distinct impact of injunctive and descriptive social norms, which have often been neglected in previous research, for instance, regarding contact tracing app use (eg, [25,26]). Thus, we provide an integrative framework to test attitudinal predictors of intention and behavior concerning contact tracing app use in the general population (Figure 3).

Figure 3. Research framework combining health behavior theories (health belief model [HBM], protection motivation theory [PMT], theory of planned behavior [TPB]) and technology acceptance (unified theory of technology acceptance and use 1 and 2 [UTAUT1 and UTAUT2]) to predict adoption intentions (and frequency) of current COVID-19 contact tracing app use. Relevant variables that are not represented in TPB or UTAUT are italicized.



Methods

Recruitment

Via a cross-sectional online survey, we explored app use and associated attitudes in a convenience sample of the German population. Participants were recruited via social media (Facebook groups, coronavirus-related websites, YouTube), press outlets (local news report, Press and Media Relations Office of the University of Greifswald), and personal communications. The survey was pretested via cognitive debriefings and pilot surveys in a small sample (n=20) for clarity, readability, accessibility, and functioning. During the pretest, we observed a duration of 10-60 minutes for survey completion (depending on literacy, familiarity with surveys, etc), which we established as a guideline for data collection. The survey was subsequently implemented using SoSci Survey (SoSci Survey GmbH) [58], over a period of 3 months (May to July 2020). This recruitment period was chosen to capture a similar time frame of about 4 weeks before and after the official launch of the German Corona Warn-App. Amidst the data collection period, the Corona Warn-App was launched by the German Ministry of Health (on June 16), so we coded participation before or after this launch to control for exposure effects.

The Corona Warn-App uses a decentralized approach to infection tracking. Via Bluetooth, it measures the distance between the smartphones of app users and exchanges temporary encrypted random codes across devices. These codes are cryptographically derived several times per hour based on a random device key. If app users test positive for the coronavirus, they can choose to pass on their keys for matching with other app users who have received the random code from this device in the given time frame. If there is a match, an algorithm determines individual risk (eg, based on estimated physical distance, timing, and duration of the meeting), and subsequently informs app users about their risk status (low or increased risk). The app is available in the App Store and the Google Play Store, and it works with multiple operating systems, starting with iOS 13.5 and Android 6.

As an incentive, 50 vouchers (€15; US $18.24; exchange rate on May 11, 2021) were randomly distributed among participants. Moreover, after study completion, participants were provided additional information on COVID-19 and tracing apps, including several hyperlinks to freely available tracing apps. The study procedure was approved by the local Ethics Committee (blinded

XSL•FO
RenderX

for peer review). A list of the items is presented in Multimedia Appendix 1.

Measurement Instruments

Sociodemographic Data

We measured individual age, gender (1=female, 2=male), number of persons in one's household, current level of education (0=lower secondary education or less; 1=upper secondary education, ie, "Abitur" or higher educational achievement), current personal income (USD values based on the exchange rate on May 11, 2021; 1=€0-€500 [US $0-$608.23], 2=€501-€1000 [US $609.44-$1216.45], 3=€1001-€1500 [US $1217.67-$1824.68], 4=€1501-€2000 [US $1825.90-$2432.91], 5=€2001-€2500 [US $2434.13-$3041.14], 6=€2501 or more [US$ 3042.35 or more]; dummy coded for the analysis with the first category as a reference category), region (0=rural, ie, up to 10,000 inhabitants; 1=urban, ie, up to 100,000 inhabitants; 2=metropolitan, ie, more than 100,000 inhabitants; dummy coded with rural as a reference category), and migration background (1=father/mother/participant born in Germany, 2=father/mother/participant born elsewhere).

Health Behavior Change

Core variables of health behavior change were assessed based on the TPB, the PMT, and the HBM (Figure 1) in line with previous studies and recommendations on scale development (eg, [53,59,60]). Mean scores were used for all scales, with higher values representing more positive attitudes.

To capture subjective norms, descriptive personal norms (eg, most people who are important to me want to use such an app) were assessed via 3 items (Cronbach α=.89) and injunctive personal norms (eg, my family expects me to use such an app) via 4 items (α=.84) on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). In addition, descriptive social norms (α=.75) (eg, how many people in your age group would like to use such an app?) and injunctive social norms (α=.97) (eg, how many people in your age group should use such an app?) were assessed with 2 items each on a scale from 0% to 100%, recoded into 10 categories (0-10, 11-20, etc) for further analysis. Perceived behavioral control was reflected by its subcomponents self-efficacy (eg, I am confident that I could use such an app) via 4 items (α=.81), and perceived controllability (eg, the decision to use such an app is up to me) via 2 items (α=.69) on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). Attitudes toward app use were assessed with 4 items (eg, good-bad, helpful–not helpful) on a 7-point semantic differential, recoded to represent positive attitudes toward app use (α=.89).

We tested the factor structure of the TPB model via a principal component analysis with varimax rotation (Multimedia Appendix 2). The resulting model (Kaiser-Meyer-Olkin [KMO]=0.80) explained about 70% of cumulative variance and achieved sufficient differentiation with item loadings above .64 on each factor. The factors largely mirrored the constructs attitudes, subjective norms, and perceived behavioral control; however, items measuring injunctive and descriptive personal norms represented one factor and were thus combined into a scale of personal norms (α=.90). Moreover, the items measuring injunctive social norms (TPB_ISN1, TPB_ISN2; see Multimedia Appendix 1) were strongly associated with the factor labelled attitudes but were kept as a separate scale for theoretical reasons.

UTAUT

To capture additional constructs described in UTAUT I, we measured performance expectancies (ie, perceived usefulness) via 6 items (eg, a personalized COVID-19 tracing app improves tracing of infection chains; α=.93) as well as the perceived costs of or barriers to app use (eg, using the app can cause problems with other apps; α=.81) on a 5-point Likert scale.

We also assessed perceived ease of use (eg, learning to use the app would be easy for me) with 4 items (α=.91) on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). To further capture the constructs specified in UTAUT2, we assessed hedonic motivation (how would you feel if you used a COVID-19 tracing app) via 3 semantic differentials (happy-sad, concerned–not concerned, satisfied-dissatisfied; α=.77), price value by eliciting the perceived material benefits of not using the app with 5 items (eg, it saves time or money; α=.88), inversed to represent positive price value, and habit (eg, using an app is something that I do often) with 6 items (α=.91) on a scale from 1 (strongly disagree) to 5 (strongly agree). All scales for the UTAUT assessment were adapted from a previous study in a German sample [61] and were in line with prior UTAUT research (eg, [48,53]). To capture experience, we also asked participants how many hours per day they spent using smartphone apps (analyzed as integers).

Like the TPB model, we examined the structure of the UTAUT model with a principal component analysis with varimax rotation (Multimedia Appendix 3). The resulting model (KMO=0.90) explained about 66% of the cumulative variance and achieved sufficient differentiation with item loadings above .58 on each factor. The factors mirrored the constructs perceived barriers, perceived ease of use, price value, and habit. However, items measuring hedonic motivation were associated with perceived usefulness but were kept as a separate scale for theoretical reasons.

As potential barriers to app use, we also included data privacy concerns (eg, a good privacy policy for mobile app users should have a clear and conspicuous disclosure) measured by the App Information Privacy Scale [62], which consists of 17 items (α=.91) rated on a 7-point Likert scale, and personalization of the app (eg, the app provides information that is exactly tailored to my needs) via 3 items (α=.90), rated on a 5-point Likert scale, to complement the analysis model.

COVID-19–Related Information

We included COVID-19–related information on risk perception, namely perceived personal susceptibility to a COVID-19 infection, as well as susceptibility of one's social surroundings and one's age group (α=.88), on a scale from 0% to 100%. Values were recoded into 10 categories (0-10, 11-20, etc) for further analysis, similar to social norms. We also measured anticipatory anxiety regarding an infection (oneself, close friends/family; α=.83) on a 5-point scale from 1 (not at all worried) to 5 (very worried), and anticipated emotion as a proxy for emotional severity in case of an infection (eg, sadness, anxiety) via 5 items (α=.81) on a 5-point Likert scale. We

XSL•FO
RenderX

assessed subjective knowledge about COVID-19 tracing apps (on a scale of 0 to 100, recoded from 1 to 10), and direct or indirect experience (ie, a close friend or family) with COVID-19, including physical (eg, respiratory illness), material (eg, unemployment), and psychological (eg, depression) consequences (0=no, 1=yes).

Finally, we assessed individual intentions to use a COVID-19 tracing app within the next 3 months via 3 items on a scale of 1 (highly unlikely) to 7 (highly likely) (eg, I plan to use a tracing app within the next 3 months; α=.99). Due to the high correlation between items, they were collapsed into a single indicator with the maximum value across all three representing individual intentions. We also measured the frequency of current tracing app use from 1 (never) to 6 (multiple times per day). To compare app users to nonusers, frequency was dichotomized into 0 (nonusers) and 1 (users) for descriptive analysis. Because the survey was implemented before the official launch of the German Corona Warn-App, response options comprised a variety of available tracing apps (eg, ito App, CoroNotes, Datenspende App) and an open-ended category (other). For analysis purposes, however, all data were collapsed to reflect (the frequency of) tracing app use via single-item measures to estimate use patterns of any kind of tracing app. To make sure that participants responded within a similar frame of reference, we provided a short explanation regarding contact tracing apps before proceeding to ask app-specific questions. This explanation described the functionality of contact tracing apps as (1) monitoring and tracking infection chains, (2) delivering immediate support and information in case of an infection or contact with an infected person, (3) and possibly providing support for persons in quarantine by monitoring health and tailoring information and preventive actions (Multimedia Appendix 4). This definition is in line with current conceptualizations of contact tracing apps [1,3,4]. However, since our study preceded the official launch of the Corona Warn-App, it was impossible to precisely describe this app and its functionalities as a point of reference. We also controlled for exposure to the app via the official launch (0=completed the survey before the official launch, 1=completed the survey after the official launch of the tracing app in Germany on June 16).

Statistical Analysis

First, we provided a descriptive analysis of the sample (including missing data) and compared current app users to nonusers via *t* tests and chi-square tests. Second, we examined bivariate associations of study variables. For bivariate correlations, we report Pearson and point biserial correlation coefficients, with boundaries of r≥0.1 (weak), r≥0.3 (moderate), and r≥0.5 (strong) as effect sizes. Third, we perform 4 block-wise hierarchical regression models of adoption intentions and frequency of current tracing app use. These models comprise the TPB (model 1), the UTAUT2 (model 2), an integrated model (model 3), and an extended model (model 4) including the additional variables threat appraisals, privacy concerns, and personalization as predictor variables. Because the latter variables are not part of the core set of constructs in either TPB or UTAUT2, we aimed to test their utility beyond the integrated framework using the "additional variables" approach (eg, [48,50]). In each regression model, the first block of predictor variables consisted of covariates, the second block of model components, and the third block of additional variables (if any). By examining the changes in explained variance for each block as well as the regression coefficient of each variable, we can consecutively examine the impact of sociodemographic data, the TPB, the UTAUT2, and additional variables on adoption intentions, and app use. We reported standardized regression coefficients (beta) and R^2 as effect sizes for each model. All analyses assumed an alpha level of .05 and were conducted with SPSS Statistics 27 (IBM Corp).

Results

Descriptive Statistics

In total, 593 persons participated in the survey; however, after excluding participants who completed the entire survey in less than 10 minutes or showed monotone response patterns for >80% of the questions, 349 participants remained (age: mean 35.62, SD 14.66 years; range 18-82 years). On average, participants completed the survey in 22.65 (SD 7.93) minutes. Descriptive statistics are presented in Table 1.

Overall, missing values were low (564/31,759, 1.8%), which the Little test [63] revealed to be missing completely at random (χ^2_{2450}=2538.4, *P*=.10). Only 3 variables had more than 5% missing values, namely income (23/349, 6.6%), region (21/349, 6.0%), and an item measuring risk perception (21/349, 6.0%). Therefore, we used complete cases for descriptive statistics and applied pairwise deletion for inferential statistics, as previous studies have shown a low probability of bias in this case [64,65]. In our sample, 19% (67/349) reported current use of a COVID-19 contact tracing app, with an average frequency of several times per week (mean 4.07, SD 1.51). A comparison of current app users and nonusers (*t* tests, chi-square tests) indicates that app use was much higher after the launch and thus exposure to the official Corona Warn-App, and seemingly equally distributed across regions. App users reported significantly more positive attitudes and fewer concerns than nonusers but also had a lower relative frequency of COVID-19 experiences. Perceived controllability of app use did not differ between participants.

XSL•FO
RenderX

Table 1. Descriptive statistics comparing current users and nonusers of a COVID-19 contact tracing app.

Characteristic	Total (N=349)	Current app users (n=67)	Nonusers (n=282)	Statistics[a]	
				Value	*P* value
Sociodemographic data					
Age, mean (SD)	35.62 (14.66)	36.62 (14.73)	35.38 (14.66)	t_{333}=0.61	.54
Gender (female), n (%)	226 (65.30)	40 (60.60)	186 (67.10)	χ^2_1=1.0	.31
Persons per household, mean (SD)	2.53 (1.58)	2.55 (1.40)	2.54 (1.63)	t_{333}=0.11	.92
Education, n (%)				χ^2_1=2.8	.10
≤Lower secondary	55 (16.50)	15 (23.40)	40 (14.90)		
Upper secondary	278 (83.50)	49 (76.60)	229 (85.10)		
Income, n (%)				χ^2_5=6.3	.28
€0-€500	66 (20.20)	12 (18.8)	54 (20.6)		
€501-€1000	70 (21.50)	8 (12.5)	62 (23.7)		
€1001-€1500	36 (11.00)	9 (14.1)	27 (10.3)		
€1501-€2000	25 (7.70)	8 (12.5)	17 (6.5)		
€2001-€2500	46 (14.10)	10 (15.6)	36 (13.7)		
€2501 or more	83 (25.50)	17 (26.6)	66 (25.2)		
Region, n (%)				χ^2_2=8.6	.01
Rural	66 (20.10)	21 (32.8)	45 (17.0)		
Urban	143 (43.60)	21 (32.8)	122 (46.2)		
Metropolitan	119 (36.30)	22 (34.4)	97 (36.7)		
Migration background[b], n (%)	76 (21.78)	16 (23.9)	60 (21.5)	χ^2_1=0.2	.67
Theory of planned behavior, mean (SD)					
Attitudes (range 1-7)	4.19 (1.65)	4.97 (1.34)	4.01 (1.67)	t_{120}=5.00	<.001
Subjective norms, mean (SD)					
Personal norms (range 1-5)	2.35 (0.87)	2.85 (0.84)	2.23 (0.83)	t_{99}=5.43	<.001
Injunctive social norms (range 1-10)	5.62 (3.52)	7.69 (2.64)	5.12 (3.53)	t_{130}=6.67	<.001
Descriptive social norms (range 1-10)	4.62 (1.92)	5.19 (1.89)	4.48 (1.90)	t_{101}=2.76	.007
Perceived behavioral control, mean (SD)					
Self-efficacy (range 1-5)	4.19 (0.85)	4.48 (0.72)	4.12 (0.87)	t_{342}=3.12	.002
Perceived controllability (range 1-5)	4.31 (0.87)	4.25 (0.95)	4.33 (0.85)	t_{342}=0.63	.53
Unified theory of acceptance and use of technology, mean (SD)					
Perceived ease of use (range 1-5)	4.04 (0.91)	4.37 (0.77)	3.95 (0.92)	t_{341}=3.45	.001
Perceived usefulness (range 1-5)	3.10 (1.08)	3.54 (0.93)	2.99 (1.09)	t_{113}=4.19	<.001
Perceived barriers (range 1-5)	2.53 (0.79)	2.26 (0.75)	2.60 (0.78)	t_{338}=3.15	.002
Hedonic motivation (range 1-5)	3.80 (1.31)	4.63 (1.18)	3.60 (1.26)	t_{337}=6.02	<.001
Price value (range 1-5)	2.97 (1.08)	3.44 (1.15)	2.86 (1.03)	t_{337}=3.95	<.001
Habit (range 1-5)	3.46 (1.08)	3.82 (0.95)	3.40 (1.07)	t_{344}=2.97	.003
Experience (ie, hours of app use per day) (range 0-12)	2.63 (1.78)	2.82 (1.34)	2.61 (1.86)	t_{122}=1.02	.31
Additional variables					

XSL•FO
RenderX

Characteristic	Total (N=349)	Current app users (n=67)	Nonusers (n=282)	Statistics[a]	
				Value	*P* value
Threat appraisals					
Perceived susceptibility (range 1-10), mean (SD)	4.14 (2.18)	4.31 (2.14)	4.10 (2.19)	t_{336}=0.68	.50
Anticipatory anxiety (range 1-5), mean (SD)	2.95 (1.03)	3.41 (0.94)	2.85 (1.02)	t_{337}=4.02	<.001
Anticipated emotion (range 1-5), mean (SD)	2.52 (0.88)	2.66 (0.91)	2.48 (0.88)	t_{337}=1.43	.16
Other variables					
Data privacy concerns (range 1-7), mean (SD)	5.71 (0.85)	5.34 (1.08)	5.79 (0.76)	t_{344}=4.00	<.001
Personalization (range 1-5), mean (SD)	2.88 (1.15)	3.36 (1.14)	2.76 (1.12)	t_{344}=3.89	<.001
Subjective knowledge about COVID-19 tracing apps (range 1-10), mean (SD)	3.69 (2.36)	5.05 (2.41)	3.37 (2.24)	t_{324}=5.27	<.001
COVID-19 experience, n (%)	161 (47.60)	23 (35.4)	138 (50.50)	χ^2_1=4.8	.03
Exposure (ie, surveyed after the Corona Warn-App launch), n (%)	135 (38.68)	54 (80.6)	81 (29.0)	χ^2_1=60.4	<.001
Adoption intentions (range 1-7), mean (SD)	3.66 (2.37)	5.78 (1.94)	3.15 (2.17)	t_{110}=9.71	<.001

[a]Two-tailed *t* test or chi-square test results.

[b]Either the participant, their mother, or their father was not born in Germany.

Bivariate Correlations

Sociodemographic data were not associated with adoption intentions, but frequency of current tracing app use was negatively correlated with education (r=–0.13, *P*=.02) and urban region (r=–0.16, *P*=.005), indicating that fewer educated participants living in metropolitan or rural areas reported more frequent tracing app use (Multimedia Appendix 5).

Regarding TPB and UTAUT constructs, Table 2 shows that adoption intentions were moderately to strongly associated with most TPB constructs (r=0.38, *P*<.001 to r=0.69, *P*<.001), except for controllability (r=0.03, *P*=.63), as well as UTAUT constructs (r=0.29, *P*<.001 to r=0.64, *P*<.001) except for experience (r=0.10, *P*=.09). Perceived costs (r=–0.35, *P*<.001) were negatively linked to intentions. Overall, frequency of current app use showed similar but weaker associations. Exposure to the app was not associated with any attitudinal variable but was associated with intentions (r=0.11, *P*=.04) and frequency of app use (r=0.39, *P*<.001).

Additionally, TPB and UTAUT constructs correlated considerably with attitudes and injunctive social norms (r=0.68, *P*<.001), attitudes and perceived usefulness (r=0.80, *P*<.001), attitudes and hedonic motivation (r=0.71, *P*<.001), with hedonic motivation and perceived usefulness (r=0.69, *P*<.001) being particularly high, underlining their conceptual similarities (cf principal component analyses in Multimedia Appendices 2 and 3). Adoption intentions and use frequency were moderately correlated (r=0.46, *P*<.001).

Among additional variables (Multimedia Appendix 5), adoption intentions were mostly strongly associated with personalization (r=0.57, *P*<.001), followed by anticipatory anxiety (r=0.41, *P*<.001), data privacy concerns (r=–0.30, *P*<.001), and anticipated emotion (r=0.24, *P*<.001). Frequency of current tracing app use had similar associations and was also positively correlated with knowledge about tracing apps (r=0.28, *P*<.001).

XSL•FO
RenderX

Table 2. Bivariate correlations between core constructs of health behavior models, technology acceptance, adoption intentions, and frequency of use of COVID-19 contact tracing apps.

Constructs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Attitudes	1															
2. Personal norms	.51[a]	1														
3. Injunctive social norms	.68[a]	.60[a]	1													
4. Descriptive social norms	.43[a]	.52[a]	.48[a]	1												
5. Self-efficacy	.37[a]	.35[a]	.33[a]	.31[a]	1											
6. Perceived controllability	.08	.06	–.04	.05	.22[a]	1										
7. Perceived usefulness	.80[a]	.48[a]	.66[a]	.41[a]	.36[a]	–.02	1									
8. Perceived barriers	–.35[a]	–.23[a]	–.26[a]	–.22[a]	–.35[a]	–.04	–.28[a]	1								
9. Perceived ease of use	.29[a]	.30[a]	.29[a]	.27[b]	.64[a]	.07	.25[a]	–.26[a]	1							
10. Hedonic motivation	.71[a]	.48[a]	.60[a]	.39[a]	.31[a]	.03	.69[a]	–.33[a]	.26[a]	1						
11. Price value	.41[a]	.26[a]	.34[a]	.26[a]	.25[a]	–.07	.36[a]	–.52[a]	.19[a]	.36[a]	1					
12. Habit	.35[a]	.19[a]	.27[a]	.23[a]	.40[a]	.04	.32[a]	–.23[a]	.37[a]	.32[a]	.25[a]	1				
13. Experience	.06	.02	.06	.14[c]	.12[c]	.02	.07	–.05	.20[b]	.10	.11	.42[a]	1			
14. Adoption intentions	.66[a]	.63[a]	.69[a]	.48[a]	.38[a]	.03	.63[a]	–.35[a]	.31[a]	.64[a]	.42[a]	.29[a]	.10	1		
15. App use frequency	.21[b]	.32[a]	.27[a]	.15[b]	.15[b]	.04	.12[c]	–.17[b]	.20[a]	.28[a]	.15[b]	.14[b]	.06	.46[a]	1	
16. Exposure	.02	.05	.05	.03	.01	.07	.00	–.04	.02	–.03	.07	–.04	–.06	.11[c]	.39[a]	1

[a]*P*<.001.

[b]*P*<.01.

[c]*P*<.05.

Hierarchical Regression Models

Results of regression models testing the TPB (model 1), the UTAUT (model 2), the integrated model (model 3), and the extended model (model 4) are presented in Table 3. In each analysis, we included several covariates in the first block (ie, age, gender, and education; number of persons per household, income, and region; as well as subjective knowledge regarding COVID-19 tracing apps, COVID-19 experience, and exposure). The selection of covariates was based on previous research on technology use, and preventive behaviors during the COVID-19 pandemic. The second block contained TPB (model 1) or UTAUT constructs (model 2) or both (model 3). The third block contained additional variables (model 4). Due to high correlations between some variables, we calculated the variance inflation factor (VIF) to test for multicollinearity [66]. Accordingly, a VIF of more than 4 is often considered an indicator of multicollinearity. The VIF was below 3.00 in model 1 and model 2, but in model 3 and 4, the VIF ranged from 1.11 (COVID-19 experience) to 4.04 (attitudes). Due to its high VIF, high bivariate correlations with other constructs, and its conceptual similarities to perceived usefulness, we excluded the variable measuring attitudes from models 3 and 4.

The TPB (model 1) explained 63% of adoption intentions, with injunctive social norms (β=.35, *P*<.001), attitudes (β=.28, *P*<.001), and personal norms having the strongest association (β=.22, *P*<.001); self-efficacy having a smaller effect (β=.08, *P*=.04); and perceived controllability being not significant (β=–.02, *P*=.56). Regarding current app use (R^2=0.33), personal norms (β=.20, *P*=.004) were connected to more frequent use, but exposure to the warning app had the strongest association (β=.31, *P*<.001). Additionally, persons from more densely populated areas reported less frequent app use.

The UTAUT (model 2) explained 56% of adoption intentions and 37% of frequent app use. Perceived usefulness (β=.32, *P*<.001) and hedonic motivation (β=.35, *P*<.001) were among the strongest predictor variables of intentions and frequency of use; however, the influence of perceived usefulness was inverted (β=–.14, *P*=.049) for frequency of use. Perceived ease of use, perceived barriers, and experience were not significant. Price value was positively associated with intentions (β=.13, *P*=.02) but not frequency, while habit was associated with frequency (β=.15, *P*=.02) but not intentions.

The integrated model (model 3) explained 67% of intentions with injunctive social norms (β=.31, *P*<.001), hedonic motivation (β=.22, *P*<.001), personal norms (β=.21, *P*<.001), and price value (β=.11, *P*=.02) remaining significant. Regarding frequency, the model explained 40% of the variance with personal norms (β=.18, *P*=.007), hedonic motivation (β=.34, *P*<.001), habit (β=.14, *P*=.03), and perceived usefulness (β=–.23, *P*=.004).

XSL•FO
RenderX

Finally, the extended model with the additional constructs privacy concerns, personalization, and threat appraisals (model 4) performed marginally better regarding intentions (R^2=0.68) and frequency of app use (R^2=0.42). Anticipatory anxiety was also significantly associated with intention (β=.09, *P*.04) and frequency (β=.15, *P*=.03). In addition, privacy concerns were associated with less frequent app use (β=–.09, *P*=.046).

To summarize, the analysis showed that the TPB as well as the UTAUT proved useful in determining intentions, and frequency of use of a COVID-19 contact tracing app, although model configurations varied greatly between both outcomes (Figure 4). While most of the constructs of the TPB were significantly associated with intentions (except for controllability), neither perceived barriers, ease of use, nor experience reached significance within the UTAUT framework. Overall, perceived usefulness, subjective norms, hedonic motivation, and anticipatory anxiety were consistently associated with tracing app use across models.

XSL•FO
RenderX

Table 3. Standardized regression weights and explained variance of hierarchical regression models predicting adoption intentions and frequency of current contact tracing app use.

Variable	TPB[a] (model 1)		UTAUT[b] (model 2)		Integrated model (model 3)		Extended model (model 4)	
	Intention	Frequency	Intention	Frequency	Intention	Frequency	Intention	Frequency
Block 1								
Covariates								
Age	.02	.03	.12	.14	.05	.10	.06	.12
Gender (ref[c]: female)	–.02	–.01	.05	.00	.00	–.03	.01	–.02
Migration background	.03	.01	.04	.01	.04	.02	.04	.03
Education (ref: ≤lower secondary)	–.02	–.10	–.02	–.08	–.02	–.07	–.02	–.08
Persons per household	–.01	–.01	–.02	.01	–.02	–.01	–.02	–.01
Income (ref: €0-€500)								
€501-€1000	.02	.02	.00	–.01	–.01	–.01	.00	.00
€1001-€1500	.03	.03	–.01	.02	.01	.02	.01	.02
€1501-€2000	–.05	–.03	–.09	–.07	–.10[d]	–.06	–.09	–.05
€2001-€2500	.05	–.01	–.03	–.05	.01	–.04	.02	–.03
€2501 or more	.06	.05	–.04	–.04	.02	–.02	.03	–.02
Region (ref: rural)								
Urban	–.06	–.18[d]	–.04	–.17[d]	–.04	–.18[e]	–.05	–.18[e]
Metropolitan	–.01	–.16[d]	–.01	–.14[d]	.01	–.14[d]	.00	–.14[d]
Subjective knowledge[f]	.06	.17[g]	.09[d]	.21[g]	.07	.20[g]	.05	.18[e]
COVID-19 experience	.01	–.05	–.02	–.08	–.01	–.07	.00	–.06
Exposure[h]	.05	.31[g]	.08	.34[g]	.06	.33[g]	.06	.33[g]
Block 2								
TPB								
Attitudes	.28[g]	.05	—[i]	—	—	—	—	—
Personal norms	.22[g]	.20[e]			.21[g]	.18[e]	.19[g]	.16[d]
Injunctive social norms	.35[g]	.13	—	—	.31[g]	.11	.29[g]	.08
Descriptive social norms	.05	–.07	—	—	.03	–.08	.04	–.07
Self-efficacy	.08[d]	.07	—	—	.03	.01	.04	.02
Controllability	–.02	–.01	—	—	.02	–.02	.02	–.02
UTAUT								
Perceived usefulness	—	—	.32[g]	–.14[d]	.11	–.23[e]	.05	–.25[e]
Perceived barriers	—	—	–.04	–.02	–.03	–.03	–.02	–.02
Perceived ease of use	—	—	.08	.08	–.01	.05	–.03	.04
Habit	—	—	.07	.15[d]	.04	.14[d]	.03	.12
Price value	—	—	.13[d]	–.05	.11[d]	–.06	.11[d]	–.05
Hedonic motivation	—	—	.35[g]	.39[g]	.22[g]	.34[g]	.21[g]	.32[g]
Experience	—	—	.00	–.01	.03	.02	.02	.01
Block 3								
Additional variables								

XSL•FO
RenderX

Variable	TPB[a] (model 1)		UTAUT[b] (model 2)		Integrated model (model 3)		Extended model (model 4)	
	Intention	Frequency	Intention	Frequency	Intention	Frequency	Intention	Frequency
Perceived susceptibility	—	—	—	—	—	—	–.08	–.05
Anticipatory anxiety	—	—	—	—	—	—	.09[d]	.15[d]
Anticipated emotion	—	—	—	—	—	—	.01	–.04
Data privacy concerns	—	—	—	—	—	—	–.04	–.09[d]
Personalization	—	—	—	—	—	—	.09	.00
Correlations (R^2)								
Block 1	.04	.23	.04	.23	.04	.23	.04	.23
Block 2	.63	.33	.56	.37	.67	.40	.67	.40
Block 3	—	—	—	—	—	—	.68	.42

[a]TPB: theory of planned behavior.

[b]UTAUT: unified theory of acceptance and use of technology.

[c]ref: reference.

[d]*P*<.05.

[e]*P*<.01.

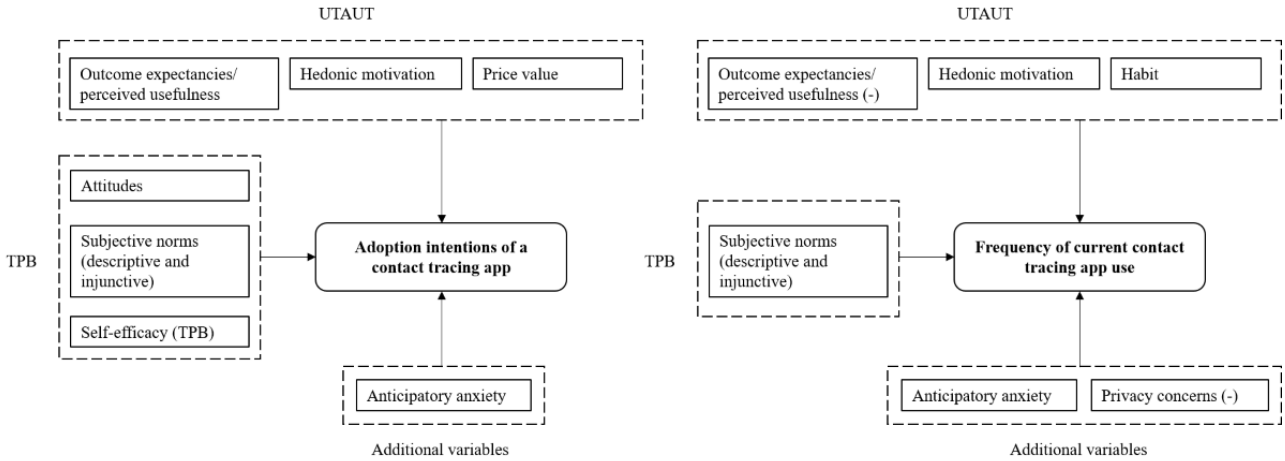
[f]Subjective knowledge about COVID-19 tracing apps.

[g]*P*<.001.

[h]Exposure (ie, surveyed after Corona Warn-App launch).

[i]Not applicable.

Figure 4. Significant predictor variables of adoption intentions and (frequency) of current COVID-19 contact tracing app use in a German community sample. Predictors are categorized according to the theory of planned behavior (TPB), the unified theory of technology acceptance and use (UTAUT), and additional variables from other health behavior models (ie, anticipatory anxiety) or previously extended UTAUT models (ie, data privacy concerns).



Discussion

Principal Results

This study aimed to investigate the utility of health behavior theories and technology acceptance models for explaining adoption intentions and current use of a contact tracing app during the COVID-19 pandemic in the German population. The TPB as well as the UTAUT explained between 56% (UTAUT) and 63% (TPB) of adoption intentions as well as 33% (TPB) and 37% (UTAUT) of current app use. Extended models including threat appraisal, privacy concerns, and personalization features of the app explained an additional 5% for both outcomes. Overall, exposure to the app (ie, study participation following an official app launch with governmental support) was strongly associated with frequent app use and was also associated with greater acceptance and more positive attitudes (Table 1). In accordance with previous research, both models have greater predictive value concerning intentions than behavior [40], which might be connected to the intention-behavior gap [38] and related factors, such as a lack of familiarity with the app, and the unprecedented and ever-changing nature of the pandemic.

Comparison With Prior Work

A closer look at the regression models also reveals that while the main components of the TPB were significantly associated with intentions—except for controllability—this was not the

XSL•FO
RenderX

case for the UTAUT model. In fact, only perceived usefulness and hedonic motivation were highly significant and consistent, while perceived barriers, experience, and ease of use were not significant. Habit and price value played a significant role for some but not all associations. Nevertheless, the large amount of explained variance points to the model's utility in explaining adoption intentions. In general, our findings support previous research on COVID-19 contact tracing apps, as we also observed strong positive associations between response efficacy (ie, perceived usefulness in UTAUT terms) and adoption intentions, but only moderate positive associations for self-efficacy [25-27]. Presumably, high ratings of perceived usefulness or positive attitudes toward the app (ie, the app being perceived as helpful; Multimedia Appendix 1) are also associated with trust in flawless performance, a low rate of false positives, and thus a sense of security. While we did not assess these constructs directly, we assume this association based on previous research on human-computer interaction [8] as well as warning systems [9]. However, future research in this context could examine the associations between trust, perceived security, and positive attitudes (eg, perceived usefulness) more closely, for instance, by using factorial surveys that vary attributes of tracing apps (eg, rate of false alarms, sensitivity) and subsequently measure attitudes [21,67].

Interestingly, the association of perceived usefulness and frequent app use was inversed, meaning that higher perceived useful was associated with lower use frequency. Given that tracing apps are designed to inform about potential COVID-19 infections and thus serious health risks, high efficacy beliefs (or perceived usefulness) would mean that more frequent app use might increase the chance of receiving health warnings which could be perceived as a negative and undesirable event. Thus, using the app less frequent could be seen as avoiding a potential threat by reducing the information flow. This risk avoidance behavior is a common phenomenon in risk research, for example, within the risk perception attitude framework [68] that connects information avoidance to high risk perceptions as well as lower self-efficacy and control beliefs. On the other hand, if perceived usefulness is low, potential health warnings issued by the app might be taken less seriously due to a higher possibility of false positives. Since only a very frequent app use guarantees preventive efficacy of the tracing app [1,3,4], these findings are concerning. We believe that future mixed methods research can elucidate the role of perceived usefulness as a potential barrier to frequent app use as well as its interaction with other predictors, such as self-efficacy and risk perception.

Furthermore, we also observed consistent significant associations between anticipatory anxiety and (frequent) app use, although the effect was small. Conceptually, these appraisals can reflect health concerns [35,60], which are associated with vigilance toward symptoms or health problems. Since a tracing app provides an additional source of information regarding a potential illness, it can help to alleviate concerns and support vigilance [1]. Similarly, frequent app use could also increase one's sense of security and positive feelings, which is mirrored by a positive association with hedonic motivation (eg, feeling satisfied, rather happy than sad, and less concerned after using the app). A similar trade-off between mHealth use, user concerns, and perceived security has already been established (eg, [1,69]). This hypothesized association between hedonic motivation, sense of security, and technology use should therefore be tested in longitudinal studies on tracing app use.

While our study underlines the importance of attitudes and perceived usefulness, we also found a very strong connection between social norms, namely personal norms (ie, referring to one's immediate surroundings), intentions, and behavior, which is contrary to prior research on tracing app use [27]. However, as previously mentioned, Kukuk [27] compiled a measure of social influence [45] that blended subjective norms, social support, and instrumental governmental support, which might be confounded by other factors such as trust in the government. Evidently, Kukuk's measure was insufficient in capturing social norms as it showed very low factorial and content validity. Additionally, in this study, injunctive social norms (ie, referring to perceived societal expectations) were also strongly connected to adoption intentions across all models. The strong impact of social norms on behavioral prevention has already been documented for previous pandemics [28,29]. With infection prevention measures (eg, social distancing) affecting the heart of social interaction, it seems reasonable to also assume a strong impact of social norms. If one's close social network is believed to support the app, one is more likely to also engage with it, leading to a social group that protects its members by using the app because of the accelerated communication in case of an infection [5]. Thus, the (frequent) use of a tracing app can be seen as an act of social responsibility. This has implications for health communication. Promoting app use as an expression of social responsibility and putting less emphasis on the perceived usefulness might be helpful in increasing public acceptance avoid potential side effects (cf the role of perceived usefulness in this study). In this sense, the approach would be quite similar to successful nudging interventions used to foster technology use [55] or vaccine uptake [56] that are based on social norms.

Moreover, two additional components of the UTAUT were significantly associated with intentions (price value) and frequency of app use (habit). In this study, price value was not solely defined by monetary restrictions [51] because the German contact tracing apps are available free of charge. Hence, we also included other limited resources (eg, time) as a potential restriction or cost. In general, the positive association with adoption intentions was in line with previous research [50,51]; however, since our operationalization of price value also included time, it is possible that the effect disappeared for frequency of use because of a foreseeable increase in time investment. Therefore, we hypothesize a trade-off between perceived costs (ie, time investment) and benefits that might differentially affect intentions and behavior. This would also explain the observed association of habit with frequency of use but not general adoption intentions. Habit describes prior experience with a technology but also the belief that using said technology is automatic [51]. Automatic behaviors are associated with less effort because it is no longer necessary to carefully and consciously plan and monitor them (eg, [53]). Since less effort equals less time spent on (consciously) using the app, habitual use may buffer the costs of more frequent app use (ie, a greater time investment). A similar trade-off is

XSL•FO
RenderX

described for effort expectancies and performance expectancies regarding technologies [50] and health behavior models (eg, [25,60]) in that a positive balance of perceived benefits over efforts predicts intentions and behaviors. However, while effort expectancies often focus on singular behaviors instead of resources, it might be beneficial to weigh perceived resource investment and behavioral efforts when predicting intentions in future studies.

Finally, we also addressed some ethical challenges that might arise in this context, namely the age-related digital divide [10,12] and privacy concerns (eg, [1,6,7]). In contrast to previous studies, we did not find any direct link between age and intentions or behavior, but older age was positively associated with privacy concerns and negatively with personalization (Multimedia Appendix 5). The negative association with personalization might reflect a lack of competence in app use in older participants, as personalizing apps requires certain skills [11,12]. Likewise, increased privacy concerns might also be associated with increased insecurity and a lack of knowledge about procedures of privacy protection implemented in the app. To concede, this study was an online study and therefore potentially excluded people who are less interested or competent in using digital technologies. However, the possibility of an indirect negative effect of age on tracing app use via privacy concerns and lower perceived benefits (ie, personalization) is alarming and has implications for health communication research. Interventional studies are needed to test whether tailored health communication surrounding benefits and privacy concerns can alleviate concerns and lead to higher acceptance and adoption of tracing apps in older participants.

Privacy concerns were negatively associated with app use, which underscores the importance of a transparent, sound, and secure data management plan when developing and promoting mHealth apps [14]. Although the effect was rather small, it is possible that participants do indeed fear misuse of health information beyond COVID-19 purposes, providing further empirical support for current publications on the promises and pitfalls of digital health solutions during the pandemic (eg, [3,6,7]). Nevertheless, further research is necessary to explore privacy concerns and their role in determining nonuse of contact tracing apps and other digital health solutions during the COVID-19 pandemic.

Limitations

The study has several limitations. First, the sample is a German convenience sample that is not representative of the population. The survey was conducted as a self-administered online survey; the reliance on self-reports might have affected our results; hence, future studies should include objective measures (eg, smartphone data on app use). In addition, although we checked the responses for monotone response patterns, we did not incorporate attention checks into the online survey. Second, regarding the cultural context of our study, we acknowledge that the technical, legal, and cultural conditions of contact tracing apps and their implementation varies greatly across countries worldwide. Since we started our survey before the official launch of the Corona Warn-App in Germany, our results are not tailored to this app but rather tracing apps in general, which might explain why we found generally lower associations for current use than for use intentions. Hence, the findings should be replicated in larger, international samples and ideally cross-cultural studies. Third, the study is cross-sectional in nature, and it was therefore not possible to longitudinally link attitudes, intentions, and behaviors as intended in the theoretical models. While strong positive exposure effects were observed, this could also point to selection bias, where proponents of the app were more strongly motivated to participate in the survey than opponents. Fourth, the measures should be tested regarding their psychometric properties. Although the instruments were based on recommendations for behavior change research, and adapted from previous research projects, their implementation in this novel context proved challenging in some areas and requires further investigation. Fourth, other constructs like trust in the government, the app provider, or health communication about COVID-19 should be included to test their impact on attitudes and the path toward adoption of the app. For example, research on the UTAUT model (eg, [48,50]) has coined the "additional variables" approach, where factors are added to the traditional UTAUT model to inspect their incremental validity. Despite these limitations, the study was based on established theories of health behavior change and technology acceptance and implemented in an ecologically valid setting. Further, it provides some insight into urgent processes of health communication and infection prevention that can support the fight against the global COVID-19 pandemic by increasing the use of contact tracing apps in the general population.

Conclusions

This study examined the utility of health behavior change and theory acceptance models in exploring intentions and use of a contact tracing app to complement preventive measures during the COVID-19 pandemic. In general, both models provide useful information for identifying core beliefs that affect intention and behavior. Among them, subjective norms, hedonic motivation, perceived usefulness, and anticipatory anxiety were particularly important across all tested models. Thus, it seems that promulgating positive social norms and addressing health concerns in health communication might be particularly beneficial to increase tracing app use in the population. The role of perceived usefulness and tracing app use needs further investigation due to its inverse associations with intentions and behavior. Moreover, privacy concerns also emerged as a barrier to app use in this context, underlining the need for more transparency and education regarding data security in the general population. Overall, the official launch of the Corona Warn-App [23] with governmental support seemed to boost awareness and app use in the population; therefore, a concerted effort is recommended when introducing a contact tracing app as an official complementary measure of infection prevention. However, data were collected between May and July 2020 amidst the first wave of the COVID-19 pandemic, so future research needs to illustrate how living under pandemic circumstances affects acceptance and use of (preventive) technology in the long term.

XSL•FO
RenderX

Acknowledgments

This study did not receive any external funding. However, we would like to thank Maxi Rahn and Johannes Fischbach for their helpful comments in designing and pretesting the study.

Authors' Contributions

All authors were responsible for study conception and design. HM, SB, and ST prepared materials and carried out data collection and analysis. ST wrote the first draft of the manuscript. All authors revised and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of items (translated from German to English).
[DOC File , 104 KB - jmir_v23i5e25447_app1.doc]

Multimedia Appendix 2
Factor loadings (above .30) of the rotated matrix of items representing the theory of planned behavior.
[DOC File , 66 KB - jmir_v23i5e25447_app2.doc]

Multimedia Appendix 3
Factor loadings (above .30) of the rotated matrix of items representing the unified theory of acceptance and use of technology.
[DOC File , 71 KB - jmir_v23i5e25447_app3.doc]

Multimedia Appendix 4
Description of a COVID-19 contact tracing app.
[DOC File , 33 KB - jmir_v23i5e25447_app4.doc]

Multimedia Appendix 5
Bivariate correlations (Pearson, point biserial) between sociodemographic data, threat appraisals, additional attitudinal variables (personalization, data privacy concerns), and adoption intentions as well as the frequency of use of COVID-19 contact tracing apps.
[DOC File , 59 KB - jmir_v23i5e25447_app5.doc]

References

1. 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 6;22(11). [doi: 10.2196/22280] [Medline: 33079693]
2. 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]
3. 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 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4]
4. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. The Lancet Digital Health 2020 Nov;2(11):e607-e621. [doi: 10.1016/s2589-7500(20)30184-9]
5. Ramakrishnan AM, Ramakrishnan AN, Lagan S, Torous J. From Symptom Tracking to Contact Tracing: A Framework to Explore and Assess COVID-19 Apps. Future Internet 2020 Sep 08;12(9):153. [doi: 10.3390/fi12090153]
6. Calvo RA, Deterding S, Ryan RM. Health surveillance during covid-19 pandemic. BMJ 2020 Apr 06;369:m1373. [doi: 10.1136/bmj.m1373] [Medline: 32253180]
7. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul 04;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
8. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 2011 Oct 16;53(5):517-527. [doi: 10.1177/0018720811417254] [Medline: 22046724]

XSL•FO
RenderX

9. Chancey ET, Bliss JP, Liechty M, Proaps AB. False Alarms vs. Misses: Subjective Trust as a Mediator between Reliability and Alarm Reaction Measures Show all authors. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Dec 20;59(1):647-651. [doi: 10.1177/1541931215591141]
10. Rudolph S. Digitale Medien, Partizipation und Ungleichheit: Eine Studie zum sozialen Gebrauch des Internets [Digital media, participation, and inequality: a study on social internet use]. Wiesbaden: Springer VS; 2019.
11. McGaughey RE, Zeltmann SM, McMurtrey ME. Motivations and obstacles to smartphone use by the elderly: developing a research framework. IJEF 2013;7(3/4):177. [doi: 10.1504/ijef.2013.058601]
12. Generali Deutschland AG. Alltag und digitale Medien [Everyday life and digital media]. In: Generali Altersstudie 2017: Wie ältere Menschen in Deutschland denken und leben. Berlin, Heidelberg: Springer; 2017:89-122.
13. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med 2020 Aug;26(8):1205-1211. [doi: 10.1038/s41591-020-0962-9] [Medline: 32546824]
14. Buck C, Burster S, Eymann T. Priming app information privacy concerns in mobile ecosystems. Bayreuther Arbeitspapiere zur Wirtschaftsinformatik, No 63. Bayreuth: Universität Bayreuth, Lehrstuhl für Wirtschaftsinformatik; 2017. URL: https://epub.uni-bayreuth.de/3419/ [accessed 2021-05-06]
15. Jalabneh R, Zehra Syed H, Pillai S, Hoque Apu E, Hussein MR, Kabir R, et al. Use of Mobile Phone Apps for Contact Tracing to Control the COVID-19 Pandemic: A Literature Review. SSRN Journal. Preprint posted online Jul 5, 2020. [doi: 10.2139/ssrn.3641961]
16. Contact tracing apps. Council of Europe. 2020. URL: https://www.coe.int/en/web/data-protection/contact-tracing-apps [accessed 2021-05-05]
17. Rowe F. Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world. Int J Inf Manage 2020 Dec;55:102178 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102178] [Medline: 32836636]
18. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A Systematic Review of Smartphone Applications Available for Corona Virus Disease 2019 (COVID19) and the Assessment of their Quality Using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
19. Martin T, Karopoulos G, Hernández-Ramos JL, Kambourakis G, Nai Fovino I. Demystifying COVID-19 Digital Contact Tracing: A Survey on Frameworks and Mobile Apps. Wireless Communications and Mobile Computing 2020 Oct 17;2020:1-29. [doi: 10.1155/2020/8851429]
20. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A Survey of COVID-19 Contact Tracing Apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
21. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. European Journal of Information Systems 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
22. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
23. Interrupt chains of infection digitally with the Corona-Warn-App. Robert Koch-Institut. 2020. URL: https://www.rki.de/EN/Content/infections/epidemiology/outbreaks/COVID-19/CWA/CWA.html [accessed 2021-05-05]
24. Übersicht zu aktuellen und früherer Zahlen und Fakten zur Corona-Warn-App. Robert Koch-Institut. 2020. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/WarnApp_KennzahlenTab.html [accessed 2021-05-05]
25. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
26. Kaspar K. Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
27. Kukuk L. Analyzing adoption of contact tracing apps using UTAUT (Essay, Bachelor). Enschede, The Netherlands: University of Twente; 2020 Presented at: 33rd Twente Student Conference on IT; July 3; Enschede, The Netherlands p. 1-8 URL: http://essay.utwente.nl/81983/
28. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
29. Weston D, Hauck K, Amlôt R. Infection prevention behaviour and infectious disease modelling: a review of the literature and recommendations for the future. BMC Public Health 2018 Mar 09;18(1):336 [FREE Full text] [doi: 10.1186/s12889-018-5223-1] [Medline: 29523125]
30. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]

XSL•FO
RenderX

31. Tomczyk S, Rahn M, Schmidt S. Social Distancing and Stigma: Association Between Compliance With Behavioral Recommendations, Risk Perception, and Stigmatizing Attitudes During the COVID-19 Outbreak. Front Psychol 2020 Aug 11;11:1821 [FREE Full text] [doi: 10.3389/fpsyg.2020.01821] [Medline: 32849073]
32. Michie S, West R. Behavioural, environmental, social, and systems interventions against covid-19. BMJ 2020 Jul 28;370:m2982. [doi: 10.1136/bmj.m2982] [Medline: 32723732]
33. Strecher V, Rosenstock I. The health belief model. In: Baum A, Newman S, Weinman J, West R, McManus CA, editors. Cambridge Handbook of Psychology, Health and Medicine. Cambridge, UK: Cambridge University Press; 1997:113-117.
34. Bandura A. Human agency in social cognitive theory. American Psychologist 1989;44(9):1175-1184. [doi: 10.1037/0003-066X.44.9.1175]
35. Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change1. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
36. 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]
37. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988 Sep 04;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
38. Sheeran P, Webb TL. The Intention-Behavior Gap. Social and Personality Psychology Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
39. Ajzen I. Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior1. J Appl Social Pyschol 2002 Apr;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]
40. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. Health Psychology Review 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
41. Lin J, Chan HC, Wei KK. Understanding competing application usage with the theory of planned behavior. J Am Soc Inf Sci 2006 Aug;57(10):1338-1349. [doi: 10.1002/asi.20453]
42. Qin L, Kim Y, Tan X. Understanding the Intention of Using Mobile Social Networking Apps Across Cultures. International Journal of Human–Computer Interaction 2018 Jan 29;34(12):1183-1193. [doi: 10.1080/10447318.2018.1428262]
43. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May 30;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
44. Clark C, Davila A, Regis M, Kraus S. Predictors of COVID-19 voluntary compliance behaviors: An international investigation. Glob Transit 2020;2:76-82 [FREE Full text] [doi: 10.1016/j.glt.2020.06.003] [Medline: 32835202]
45. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
46. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
47. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
48. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. Inf Syst Front 2017 Jun 8;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]
49. Nunes A, Limpo T, Castro SL. Acceptance of Mobile Health Applications: Examining Key Determinants and Moderators. Front Psychol 2019 Dec 10;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
50. Venkatesh V, Thong J, Xu X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
51. Venkatesh V, Thong J, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
52. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
53. Sun Y, Wang N, Guo X, Peng Z. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. J Electron Commer Res 2013;14(2):183-200 [FREE Full text]
54. Cialdini R, Trost M. Social influence: social norms, conformity and compliance. In: Gilbert DT, Fiske ST, Lindzey G, editors. The Handbook of Social Psychology, Vol 1-2, 4th Ed. New York City, NY: McGraw-Hill; 1998:151-192.
55. Schneider D, Klumpe J, Adam M, Benlian A. Nudging users into digital service solutions. Electron Markets 2019 Nov 28;30(4):863-881. [doi: 10.1007/s12525-019-00373-8]
56. Korn L, Betsch C, Böhm R, Meier NW. Social nudging: The effect of social feedback interventions on vaccine uptake. Health Psychol 2018 Nov;37(11):1045-1054. [doi: 10.1037/hea0000668] [Medline: 30221969]
57. Mols F, Haslam SA, Jetten J, Steffens NK. Why a nudge is not enough: A social identity critique of governance by stealth. Eur J Polit Res 2014 Nov 21;54(1):81-98. [doi: 10.1111/1475-6765.12073]
58. Leiner DJ. SoSci Survey (Version 3.2.27). 2019. URL: https://www.soscisurvey.de/en/index [accessed 2021-05-05]

XSL•FO
RenderX

59. Ajzen I. Constructing a TpB questionnaire: conceptual and methodological considerations. CiteSeerX. 2006. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.601.956&rep=rep1&type=pdf [accessed 2021-05-05]
60. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice, 4th Ed. San Francisco, CA: Jossey-Bass; 2008:45-65.
61. Fischbach J. Determinanten der Technologie- und Prozessakzeptanz im Kontext kooperativer Arbeit. Z. Arb. Wiss 2019 Jan 24;73(1):35-44. [doi: 10.1007/s41449-019-00146-1]
62. Buck C, Burster S. App information privacy concerns. 2017 Presented at: 23th Americas Conference on Information Systems (AMCIS); August 10-12; Boston, MA.
63. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
64. Dong Y, Peng CJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
65. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999 Mar 02;8(1):3-15. [doi: 10.1177/096228029900800102] [Medline: 10347857]
66. Thompson CG, Kim RS, Aloe AM, Becker BJ. Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. Basic and Applied Social Psychology 2017 Feb;39(2):81-90. [doi: 10.1080/01973533.2016.1277529]
67. Jasso G. Factorial Survey Methods for Studying Beliefs and Judgments. Sociological Methods & Research 2016 Jun 30;34(3):334-423. [doi: 10.1177/0049124105283121]
68. Turner MM, Rimal RN, Morrison D, Kim H. The Role of Anxiety in Seeking and Retaining Risk Information: Testing the Risk Perception Attitude Framework in Two Studies. Human Comm Res 2006 Apr;32(2):130-156. [doi: 10.1111/j.1468-2958.2006.00006.x]
69. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WS, et al. Consumer Attitudes and Perceptions on mHealth Privacy and Security: Findings From a Mixed-Methods Study. J Health Commun 2015 Apr;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]

Abbreviations

HBM: health belief model
KMO: Kaiser-Meyer-Olkin
mHealth: mobile health
PMT: protection motivation theory
TPB: theory of planned behavior
UTAUT: unified theory of acceptance and use of technology
VIF: variance inflation factor

Edited by G Eysenbach; submitted 02.11.20; peer-reviewed by F Pecoraro, D Fedele, L König; comments to author 08.12.20; revised version received 18.12.20; accepted 11.04.21; published 19.05.21.

Please cite as:
Tomczyk S, Barth S, Schmidt S, Muehlan H
Utilizing Health Behavior Change and Technology Acceptance Models to Predict the Adoption of COVID-19 Contact Tracing Apps: Cross-sectional Survey Study
J Med Internet Res 2021;23(5):e25447
URL: https://www.jmir.org/2021/5/e25447
doi:10.2196/25447
PMID:33882016

©Samuel Tomczyk, Simon Barth, Silke Schmidt, Holger Muehlan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis

Stephen Wai Hang Kwok[1], PhD, RN; Sai Kumar Vadde[2], BSc; Guanjin Wang[2], PhD

[1]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

[2]Discipline of Information Technology, Media and Communications, Murdoch University, Perth, Australia

Corresponding Author:
Guanjin Wang, PhD
Discipline of Information Technology, Media and Communications
Murdoch University
90 South St
Perth, 6150
Australia
Phone: 61 0893607351
Email: Guanjin.Wang@murdoch.edu.au

Abstract

Background: COVID-19 is one of the greatest threats to human beings in terms of health care, economy, and society in recent history. Up to this moment, there have been no signs of remission, and there is no proven effective cure. Vaccination is the primary biomedical preventive measure against the novel coronavirus. However, public bias or sentiments, as reflected on social media, may have a significant impact on the progression toward achieving herd immunity.

Objective: This study aimed to use machine learning methods to extract topics and sentiments relating to COVID-19 vaccination on Twitter.

Methods: We collected 31,100 English tweets containing COVID-19 vaccine–related keywords between January and October 2020 from Australian Twitter users. Specifically, we analyzed tweets by visualizing high-frequency word clouds and correlations between word tokens. We built a latent Dirichlet allocation (LDA) topic model to identify commonly discussed topics in a large sample of tweets. We also performed sentiment analysis to understand the overall sentiments and emotions related to COVID-19 vaccination in Australia.

Results: Our analysis identified 3 LDA topics: (1) attitudes toward COVID-19 and its vaccination, (2) advocating infection control measures against COVID-19, and (3) misconceptions and complaints about COVID-19 control. Nearly two-thirds of the sentiments of all tweets expressed a positive public opinion about the COVID-19 vaccine; around one-third were negative. Among the 8 basic emotions, trust and anticipation were the two prominent positive emotions observed in the tweets, while fear was the top negative emotion.

Conclusions: Our findings indicate that some Twitter users in Australia supported infection control measures against COVID-19 and refuted misinformation. However, those who underestimated the risks and severity of COVID-19 may have rationalized their position on COVID-19 vaccination with conspiracy theories. We also noticed that the level of positive sentiment among the public may not be sufficient to increase vaccination coverage to a level high enough to achieve vaccination-induced herd immunity. Governments should explore public opinion and sentiments toward COVID-19 and COVID-19 vaccination, and implement an effective vaccination promotion scheme in addition to supporting the development and clinical administration of COVID-19 vaccines.

(J Med Internet Res 2021;23(5):e26953) doi:10.2196/26953

KEYWORDS

COVID-19; vaccination; public topics; public sentiments; Twitter; infodemiology; infoveillance; social listening; infodemic; social media; natural language processing; machine learning; latent Dirichlet allocation

XSL•FO
RenderX

Introduction

The COVID-19 Pandemic

COVID-19 is an infectious disease caused by the novel coronavirus SARS-CoV-2, which was first identified in Wuhan, China, in December 2019 [1]. As of early January 2021, the cumulative number of confirmed cases was 83,862,300, while the number of deaths was 1,837,253, affecting 222 countries or regions globally [2]. In Australia, the total number of confirmed cases was 28,483, and the number of deaths was 909 in early January 2021 [3]. Both the incidence and prevalence have been rising globally, although these rates differ across countries [4]. In 2020, the pandemic had significant negative impacts on individuals, governments, and the global economy [5,6].

Patients with COVID-19 could experience either no symptoms, common signs and symptoms of infection, or respiratory distress, or die from the disease. The proportion of asymptomatic patients was estimated at 16%, with the proportion in children being nearly double that of adults [7,8]. However, over 80% of those who were asymptomatic had either unilateral or bilateral pulmonary involvement in computerized tomography scans [8]. Among those who were symptomatic, fever, cough, and fatigue were the most common symptoms [9,10]. Five percent of patients with COVID-19 developed acute respiratory distress syndrome [11]. Among them, the death rate ranged between 13% and 69% across countries [12].

The virus could be transmitted through close contact, or even droplets, between individuals, where the mucous membranes of healthy individuals are exposed to secretions produced by the carriers [13]. The reproductive number (R_0) of COVID-19 was approximately 3 but varies from 2 to 7 across countries [14,15]. This means one carrier could infect 3 individuals on average. Under public infection control measures, social distancing does not seem applicable to family households where the risk of transmission is high. A meta-analysis of 24 studies found that the intrafamily transmission rate of SARS-CoV-2 was higher than the transmission rate of severe acute respiratory syndrome coronavirus (SARS-CoV) or Middle East respiratory syndrome coronavirus (MERS-CoV) in households [16], which may contain vulnerable groups such as the elderly, those who are immunocompromised, or have chronic diseases.

Background on Vaccination

Briefly, the purpose of vaccination is to allow the immune system to memorize the features of the targeted pathogen and be able to initiate an immune response that is fast and strong enough to defeat the live pathogen in the future. Over 115 vaccines for COVID-19 are undergoing investigation and trials, and most of them target the spike protein of SARS-CoV-2 [17]. The development of a vaccine usually takes years. The relatively fast development of the COVID-19 vaccine could be ascribed to previous work on vaccines for SARS-CoV, which is 80% similar to SARS-CoV-2, as well as the immense and urgent need for vaccination [18].

Vaccination that is evidence-based and officially approved by health authorities is generally safe. The adverse effects, as well as their incidence rates, vary across types of vaccines. Previous studies have reported the incidence rates of severe adverse reactions in general populations after receiving vaccines. For example, the incident rate of febrile seizures after receiving the measles, mumps, and rubella (MMR) and varicella vaccine was 8.5 per 10,000 doses [19]. The rates attributable to influenza vaccines or 13-valent pneumococcal conjugate vaccines (PCV13) were 13 to 45 per 100,000 doses [20]. On the other hand, the incident rate of thrombocytopenic purpura after MMR injection was 1 per 20,000 doses [19]. Moreover, the incidence rates of some rare diseases such as intussusception after rotavirus vaccine injection ranged from 1 to 5 per 100,000 doses [20]. There was insufficient evidence to conclude that vaccination was the direct cause of the severe adverse effects compared with the vast majority of those who benefited from vaccinations.

Vaccination is a collective strategy that needs a high proportion of the population to be vaccinated in order to generate a protective effect. The proportion is calculated as $(R_0–1)/R_0$ [21]. If one patient could infect 3 individuals, then the proportion of the population that needs to be vaccinated would be two-thirds. This two-thirds should comprise individuals who have normally functioning immune systems. Those who are immunocompromised are contraindicated to certain types of vaccines such as live vaccines because of poor responses or severe adverse reactions [22,23]. Severe allergic reaction to a vaccine is a contraindication, although the risk is as small as 1 per 1,000,000 doses [19]. Hence, the higher the proportion of those who have normal immune systems receiving vaccinations, the better for achieving herd immunity to protect oneself and others.

Exploring Public Opinion on the COVID-19 Vaccine

In the last two decades, a prominent antivaccination movement has risen, resulting in a decline in MMR vaccination coverage and a rise in measles outbreaks in the United States, the United Kingdom, and certain major European countries [24]. A case study, which proposed an association between the MMR vaccine and autism [25], although disproven by several studies in subsequent years [26-31], fueled the antivaccine movement, and then was retracted [32]. Nevertheless, the adverse factors promoting antivaccination might be ignoring high-level evidence such as the results of randomized controlled trials of vaccines [33-35] as well as a selective adoption of unverified information by the public.

Social media has become a frequently used platform to disseminate both authorized information and misinformation. Authorized sources such as the World Health Organization [36], the US Centers for Disease Control and Prevention [37], the US Food and Drug Administration [38], and the UK Department of Health and Social Care [39] are available online. However, previous studies showed that around 30% to 60% of the information related to vaccination on social media were antivaccine content [24]. In websites that provided vaccine-related information, over 50% contained inaccurate information [40]. Although antivaxxers proposed different rationales to oppose vaccination [41], the fact is that only vaccination has a history of successfully eradicating viral diseases such as smallpox [42].

XSL•FO
RenderX

As several COVID-19 vaccine trials are progressing to or have nearly completed phase 3 in the second half of 2020, it is expected that vaccines will be made available to the public by 2021 [43,44]. In Australia, Dodd et al [45] conducted an online survey of 4362 adults in mid-April 2020 about 1 month after lockdown measures had been imposed. They found that 86% of the sample claimed that they would get the COVID-19 vaccine when available. At that time, 65% to 75% of the respondents were confident in the federal and state governments' responses. On August 19, the Australian prime minister [46] announced that the government had made an agreement with AstraZeneca: if its COVID-19 vaccine is proven to be safe and effective, Australia could manufacture it and make it free for the public. Later, the University of Oxford and AstraZeneca [47] and Johnson & Johnson [48] paused their vaccine trials in mid-September and mid-October 2020, respectively, to investigate adverse reactions among participants during the trials, which were resumed after investigations.

Significant health care–related events, such as news about vaccine efficacy [49], disease outbreak [50], or legislative decree of mandatory vaccinations [51], were found to trigger public discussions on social media. However, negative news about the vaccine, as well as antivaccine sentiment, could be hurdles to achieving vaccination-induced herd immunity. For example, information associated with the adverse effects of vaccinations were commonly manipulated by antivaxxers to fuel their movements [52]. They had even started using conspiracy theories against developing COVID-19 vaccines even before development had begun [53-55]. Therefore, online public opinion and sentiments around COVID-19 vaccination need to be explored and reviewed to promote public vaccination schemes based on factors affecting vaccination acceptance.

This study aimed to explore major topics and sentiments of tweets about COVID-19 vaccination among Twitter users in Australia. Findings from this study could help governments and health agencies plan, modify, and implement a timely promotion of vaccination to achieve vaccination-induced herd immunity.

Methods

Data Collection

Twitter, one of the world's major social media platforms, with 187 million daily active users as of the third quarter of 2020 [56], was chosen as the data source. Twitter is a common source of text for sentiment analysis [57,58] and analysis of sentiments toward vaccinations [59,60]. We used the R library package *rtweet* [61] to access the Twitter premium API (application programming interface) service and collect COVID-19 vaccine–related tweets posted between January 22 and October 20, 2020. Retweets, non-English tweets, and tweets with a geolocation outside Australia were excluded. The search terms "vacc OR vax OR vaccine OR vaccination" AND "corona OR covid" were used to search target tweets. Boolean operators "AND" and "OR" guaranteed that tweets that contained words belonging to the root of "vaccine" as well as the root of either "coronavirus" or "COVID" could be searched. As a result, 31,100 tweets were collected and used in this study. The number of tweets collected from January 22 to October 20, 2020, are shown in Multimedia Appendix 1.

Data Preprocessing

The R library packages of *qdapRegex* [62] and *tm* [63,64] were used for the preprocessing of text. The procedures included (1) removal of non-English words or common words that do not provide insights into a specific topic (eg, stop words); (2) case folding, which changes words into lower case for stemming; and (3) stemming of inflected words into roots, followed by stem completion to return complete words (tokens) for the results visualizations. The custom stop words removed were "amp" (ampersands) and the inflected words derived from "vaccine," "coronavirus," and "COVID." In addition to that, all stop words with reference to those in the package *tm*, Python libraries *spaCy* [65] and *gensim* [66], as well as stop words suggested by Sedgewick and Wayne [67] and the SAO/NASA (Smithsonian Astrophysical Observatory/National Aeronautics and Space Administration) Astrophysics Data System [68], were also removed in the corpus. Stop words in Python libraries and in other aforementioned sources were extracted and assigned to an R object for the ease of process in R. In addition, the dictionary used for stem completion was a corpus saved before the stemming procedure.

Associations Between Word Tokens

The word tokens were sorted by their counts in the corpus and plotted against their counts as shown in Multimedia Appendix 2. It was observed that the inflection point of the concave-up, decreasing curve was located at approximately 250 counts. Thus, word tokens having counts greater than 250 were included in pairwise correlation tests. The R library package *widyr* [69] was used to compute the correlations between word tokens. Then, the word pairs with Pearson correlation coefficients larger than 0.1 were plotted in a network graph. Coefficients smaller than 0.1 were considered negligible [70,71]. On the other hand, word pairs were also sorted by their counts and plotted against the counts as shown in Multimedia Appendix 3. Word pairs having counts larger than 150 were plotted in another network graph. The cutoff of 150 was adopted so that major clusters of word pairs with higher counts could be identified in the network without overly suppressing other pairs with significantly lower counts.

Latent Dirichlet Allocation Tuning and Model Building

Latent Dirichlet allocation (LDA) [72] is an unsupervised machine learning method that allows observations such as words or documents in a corpus to be explained by latent groups such as topics. LDA has been used in topic modeling of public opinions on certain vaccinations for human papillomavirus (HPV) [73] and influenza virus [74]. However, LDA topic modeling on COVID-19 vaccination was yet to be done. The corpus preprocessed was converted into a document-term matrix, and then terms that were sparse by less than 99.9% were retained for LDA modeling. The R library package *ldatuning* [75] was used to estimate the optimal number of topics in the LDA model. Four different metrics were computed in a range of topics (2-50) to identify the optimal number (Multimedia Appendix 4). The lower the metrics of

XSL•FO
RenderX

"Arun2010" [76] and "CaoJuan2009" [77], and the higher the metrics of "Griffiths2004" [78] and "Deveaud2014" [79], indicated a better number of topics to fit the LDA model. In this study, the metric of "Deveaud2014" reached its highest level and the metric of "CaoJuan2009" reached one of the lowest levels at 3 topics that were adopted as the number of topics for LDA modeling. Another R library package *topicmodels* [80] was used to estimate the two posterior Dirichlet distributions—theta distribution over the 3 topics within each tweet and beta distribution over all words within each topic. Only the top 100 words with the highest beta values were visualized using a word cloud for each topic. A larger font size and a higher level of opacity were used to indicate words with higher beta values. In each topic, the top 20 tweets, except those from news sources, with the highest theta values, which were also larger than those of the other two topics for each tweet, were reported.

Sentiment Analysis

The R library package *syuzhet* [81], which applies Stanford's CoreNLP [82] on text against an emotion dictionary, was used to score each tweet based on the 2 sentiments and 8 emotions defined in the Canadian National Research Council's Word-Emotion Association Lexicon [83,84]. There were 10 categories for scoring a tweet. The 2 sentiments were negative and positive, while the 8 emotions were anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. The polarity of a tweet could be positive or negative, whereas emotion recognition aimed to identify the emotions that a tweet carried. If a tweet was associated with a particular emotion or sentiment, it would score points that reflect the degree of valence with respect to that category. Otherwise, it would have no score for that category.

Results

Overview

We first analyzed the preprocessed tweets by visualizing the word tokens with a count of >250 in the corpus as shown in the word cloud in Multimedia Appendix 5. The larger the word font size in the cloud, the higher the number of counts in the corpus. The top 10 high-frequency words were "trials," "australia," "virus," "news," "developers," "flu," "people," "years," "world," and "testing." Following that, other frequently used words included: "research," "working," "timeline," "immune," "australian," "effects," "russian," "health," "human," and "government." Based on the descriptive statistics of word counts, news about the pandemic, seasonal flu, and vaccine trials were major discussion topics among Australians. Other topics such as the effects of infection control strategies and immunity, the situation overseas, and the government's responses were also relatively prominent.

Figure 1 shows the network of word pairs with counts above 150 in the corpus. The word tokens linked with edges, where thicker and more opaque lines indicate a higher number of counts. From the graph, a group of words that were frequently used together were "trials," "human," "clinical," "news," and "australia." Moreover, the word "trials" was linked to a number of word tokens such as "phase," "australia," "testing," "volunteers," and "university"; the latter was linked to "oxford" and "queensland." Another cluster of words that were commonly used together included "flu," "years," "virus," and "people." Bigrams such as "herd" and "immune" had some associations with "flu" and "virus." There were a few word pairs, such as "antivax" and "vaxxers," which were not connected to the main network and had a relatively small number of counts at the periphery of the graph.

We further examined the correlations between word tokens. The network of correlations (r>0.10) between word tokens with a count above 250 in the corpus is visualized in Figure 2, where the edges with a larger width and higher opacity indicate stronger correlations between word tokens. A major network of words consisted of keywords associated with the development and clinical trial of vaccines such as "trials," "clinical," "human," "phase," "volunteers," "participant," "astrazeneca," "university," "queensland," and "oxford." Another noteworthy major word network was composed of keywords that were related to the Australian government's partnership with vaccine manufacturers in providing doses for the public: "deal," "federal," "government," "scotty," "morrison," "millions," and "doses." On the other hand, "flu" was the center of another cluster associated with "influenza," "deaths," "rates," "vax," and "shot." Some word pairs like "common" and "cold," "herd" and "immune," and "antivax" and "vaxxers" had distal associations with the main network. The pair "antivax" and "vaxxers" had some associations with "conspiracies" and "vax" linking with "flu" and "understand," which in turn correlated with "science" and "shared." Furthermore, "social" and "distancing" had a strong correlation, but this bigram, along with a few words that had some associations with them, did not link with the larger network of word tokens. Other similarly independent bigrams included "fast track" and "big pharma."

We built a 3-topic LDA model and visualized the top 100 probability (beta) distributions of words for each topic in word clouds (Multimedia Appendix 6). The beta values are reported in Multimedia Appendix 7, and the top 20 probability (theta) distributions of topics in the tweet samples are shown in Multimedia Appendices 8-10. Three topic themes were synthesized from the word clouds and tweets extracted.

XSL•FO
RenderX

Figure 1. Network of word pairs with counts above 150 in the corpus.

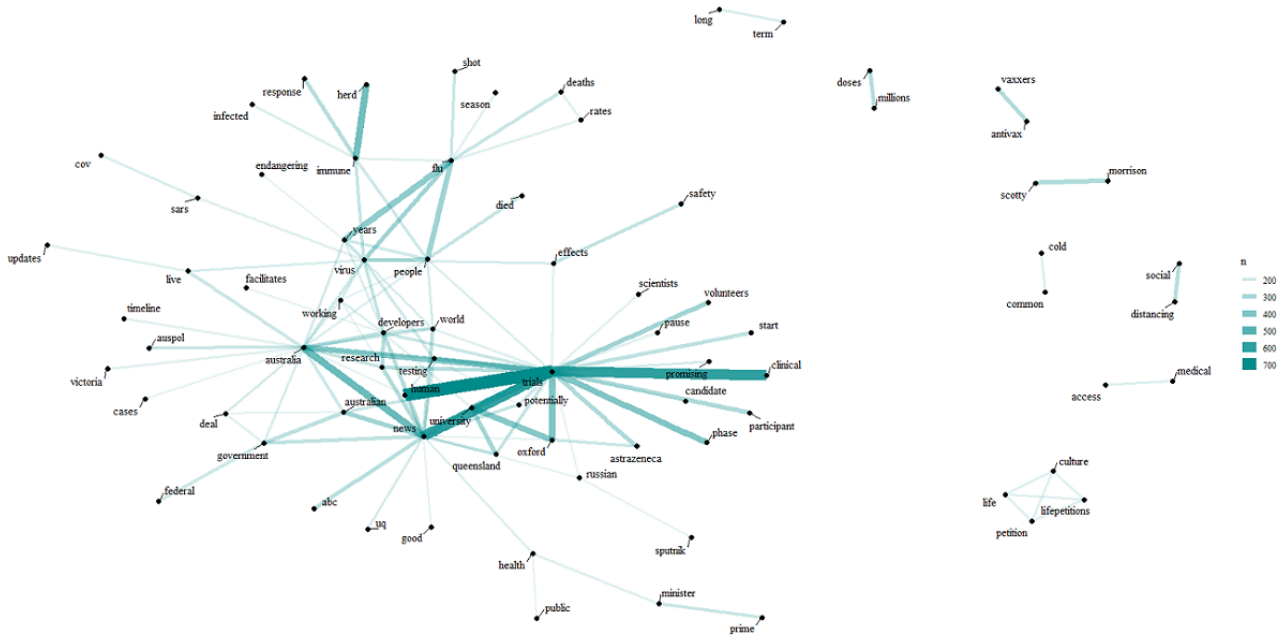
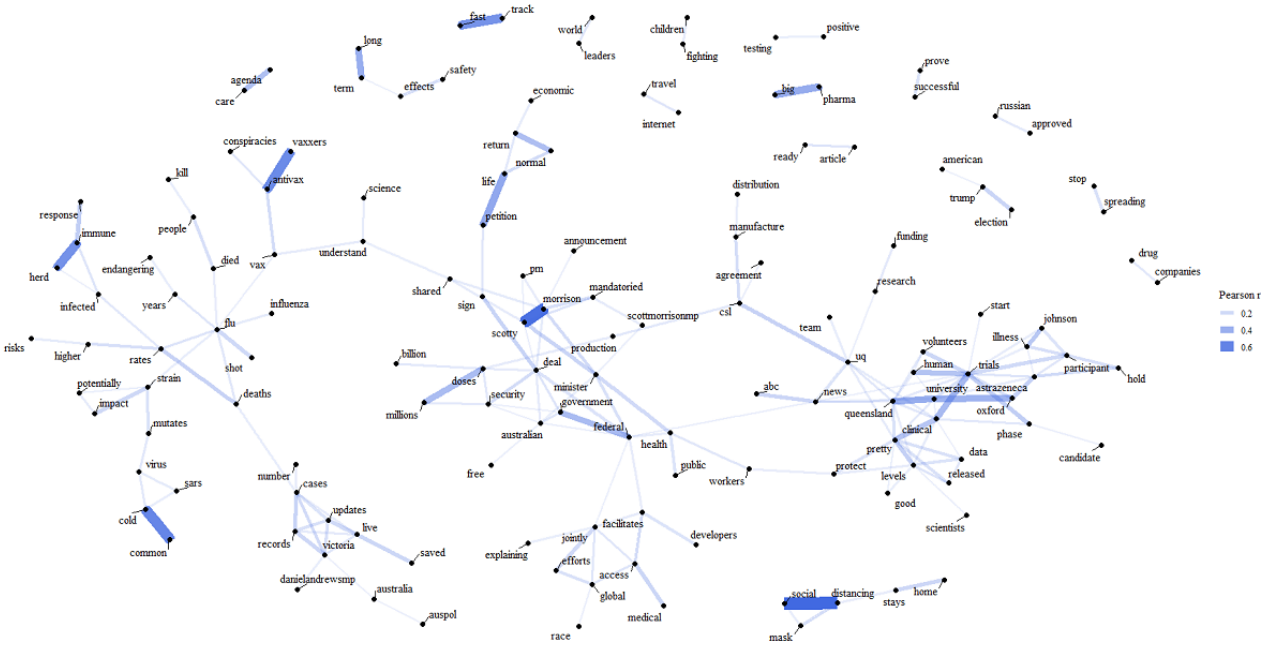


Figure 2. Network of correlations (r>0.10) between word tokens with counts above 250 in the corpus.



Topic 1: Attitudes Toward COVID-19 and COVID-19 Vaccination

The latent topic 1 centered on the public's attitudes or actions toward COVID-19 vaccination, which were associated with personal values, theories, information received, or personal experiences. Vaccine supporters accepted COVID-19 vaccination because they considered that measures should be taken to cope with the rising number of infections, deaths, health care burden, and costs due to COVID-19. They scorned those who pretended to be experts or posted misinformation such as claiming that deaths from COVID-19 were attributable to other diseases. In addition, they also supported public vaccinations by taking actions such as seeking funding sources and media to promote vaccine trials. Those who worried about the COVID-19 vaccine were skeptical about conspiracy theories such as the "mark of the beast" and microchips in vaccines. The sudden pause of vaccine trials also triggered worries among users about the safety of vaccination. Some Twitter users claimed that they would not get vaccinated because of previous experience with vaccination-related adverse effects. Nonetheless, stock prices increased when positive news about vaccine development were released. Other Twitter users disregarded COVID-19, expressing that COVID-19 had a much lower death rate than the flu, thus making it insignificant for vaccination, which they deemed would only benefit pharmaceutical firms

XSL•FO
RenderX

or be politicized. Moreover, implementing lockdowns before mass vaccination was not considered efficient in the long run. Users also thought that COVID-19 should not deserve more attention than other global problems such as climate change, aged care, or other diseases.

Topic 2: Advocating for Infection Control Measures Against COVID-19

The latent topic 2 indicated that some Twitter users were positive about the development of COVID-19 vaccines and antivirals and recognized the need for these products. Meanwhile, they also advocated following infection control measures and disproved misinformation or conspiracy theories. Some Twitter users rebutted tweets that may have been posted by antivaxxers or conspiracy theorists. For example, these users refuted skepticism over the safety of the rapidly produced vaccines, false claims about the association between the flu vaccine and COVID-19 infections and deaths, and inaccurate beliefs about vaccination coverage for achieving herd immunity differing across diseases. Some of their tweets emphasized the rising number of deaths related to COVID-19 within a rather short period compared with other pandemics in the past. They argued that although there were deaths caused by the flu, there were drugs, vaccines, and promotion campaigns targeting the flu. In comparison, deaths from COVID-19 were soaring, and even worse than the flu, without mass vaccinations or antivirals. However, COVID-19 deaths could have been preventable. With previous experiences in developing vaccines for other coronaviruses such as MERS-CoV, users believed that the COVID-19 vaccine could be successfully developed to protect vulnerable groups like patients. They believed that everyone was susceptible to COVID-19 after contracting the coronavirus without vaccination. In the future, antivirals could also be developed. Beyond vaccines and drugs, they thought physical measures such as wearing masks and social distancing should be followed, particularly at a time when mass vaccination and antivirals are not yet available.

Topic 3: Misconceptions and Complaints About COVID-19 Control

The latent topic 3 generally showed the baseless claims and conspiracy theories that antivaxxers held against the COVID-19 vaccine as well as complaints and helplessness about testing and lockdown measures, which would likely end with vaccination-induced herd immunity. Some Twitter users made claims that were unfounded or based on conspiracy theories against the COVID-19 vaccine. For example, one concluded that Australia suggested using a vaccine that had never been tested or certified to fight the virus. Some others believed that hydroxychloroquine was an effective treatment; hence, banning its use was viewed as a politicized action. Users also thought that those rejecting hydroxychloroquine should take vaccines from Bill Gates, who was falsely accused of planning to implant microchips into human bodies via vaccinations. However, other Twitter users pointed out the limitations of vaccinations such as their inability to prevent viral transmissions or treat COVID-19 and its complications. Even if vaccines are available, a high number of doses globally and tests for the virus or even antibodies are required if COVID-19 is not eradicated. Some complained that the tests led to an increase in known positive cases and in turn a prolonged lockdown, making the situation helpless without the availability of a vaccine. On the other hand, provaxxers celebrated the success in vaccine development. They criticized antivaxxers for not believing in science and accepting vaccination, as well as for disregarding the serious consequences of COVID-19 and for suggesting natural herd immunity, which would be catastrophic. For example, allowing the rampant spread of the coronavirus would lead to health care system breakdown and loss of life.

Figure 3 shows the change in sentiment scores of all tweets between January and October 2020. In each tweet, there could be both positive and negative sentiment with valences in opposite directions. Figure 3 shows that the scores increased gradually between January and March 2020. The higher the sentiment score, regardless of direction, the likelier the tweet will have stronger sentiments. However, most tweets expressed positive sentiment (score=62,498, 67%) rather than a negative one (score=27,622, 30%), while 940 (3%) tweets were neutral.

Figure 4 shows the emotion scores with respect to anticipation, joy, surprise, and trust in all tweets. The scores also rose in the first quarter of 2020. Approximately 45% of the scores were associated with these 4 emotions. Specifically, the emotion components were trust (score=22,436, 17%) and anticipation (score=19,278, 14%). Some tweets scored for surprise (score=7865, 6%) and joy (score=10,296, 8%).

Figure 5 shows the scores of negative emotions such as anger, disgust, fear, and sadness for all tweets. The scores increased in the first 3 months of 2020; approximately one-third of the scores were associated with these negative emotions. Among them, fear was the most significant one (score=18,449, 14%). Other emotions included sadness (score=11,082, 8%), anger (score=9091, 7%), and disgust (score=6337, 5%). On the other hand, nearly 22% (n=6994) of the tweets were emotionally neutral.

XSL•FO
RenderX

Figure 3. Distributions of sentiment valences between January and October 2020.

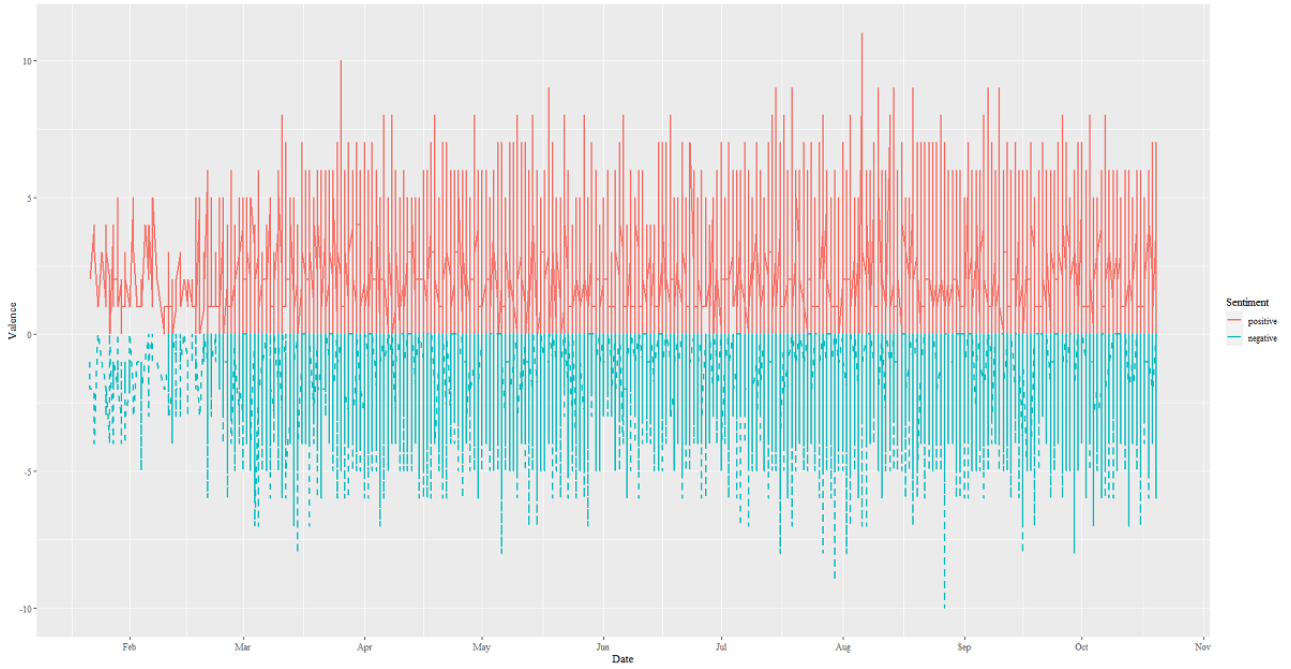
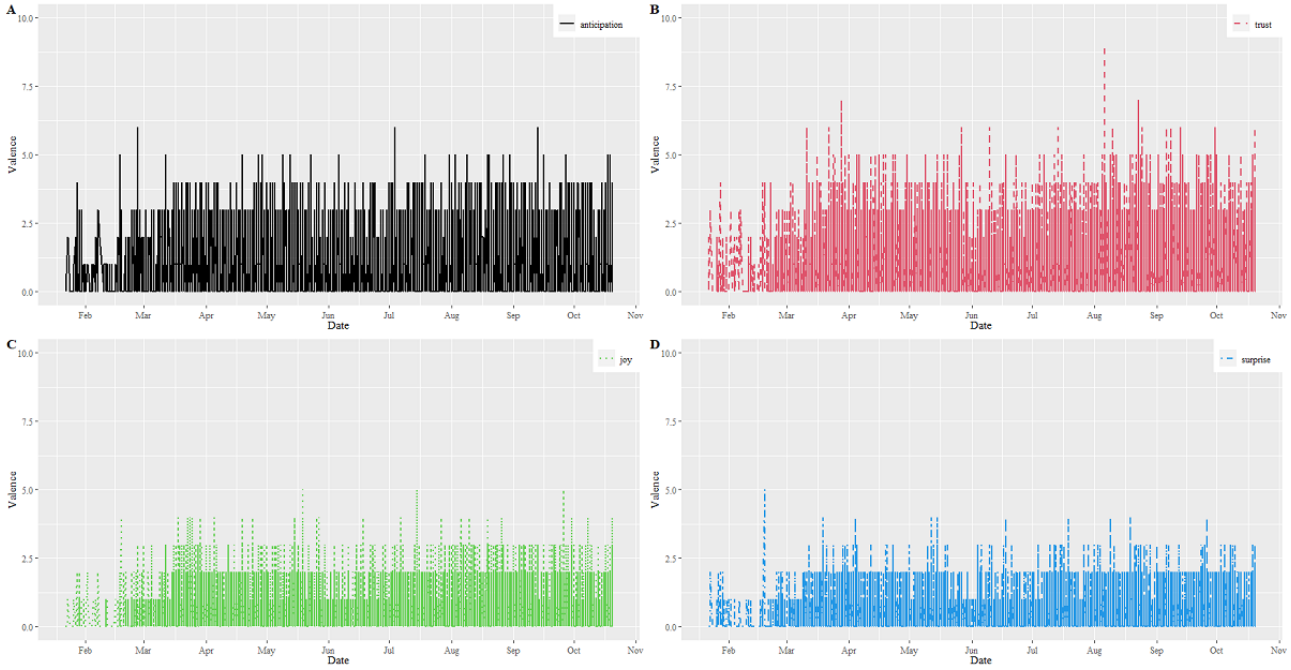
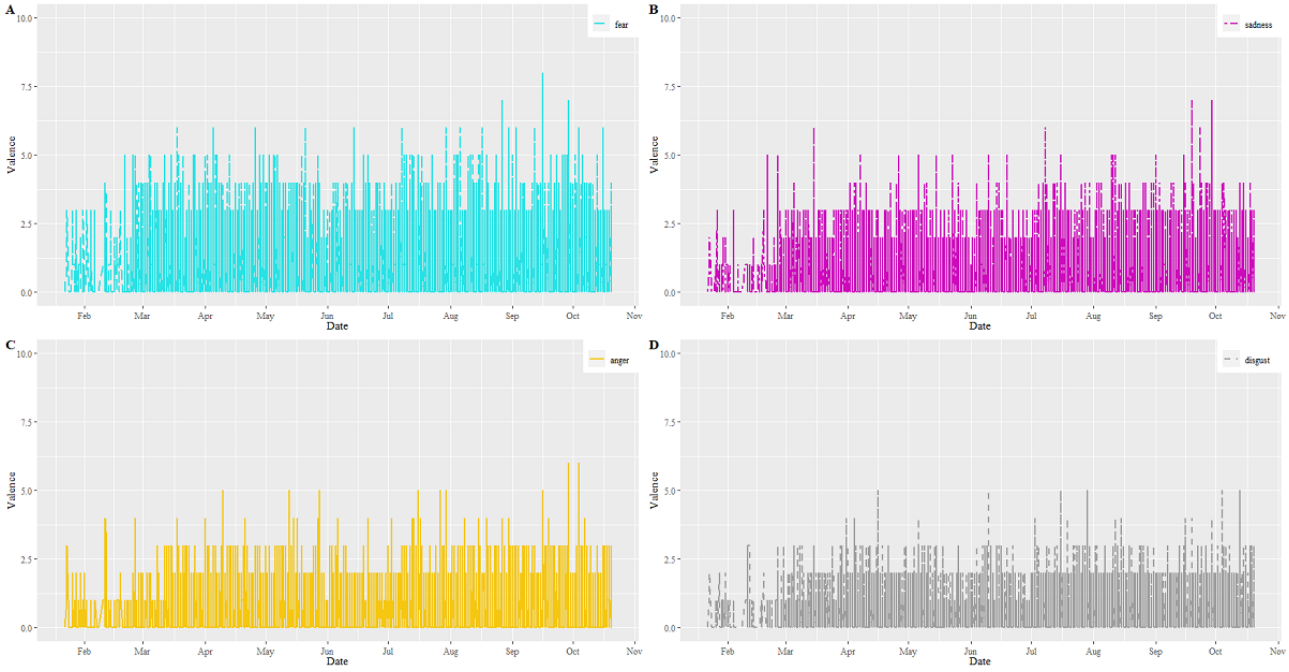


Figure 4. Distributions of emotion valences for (A) anticipation, (B) trust, (C) joy, and (D) surprise between January and October 2020.



XSL•FO
RenderX

Figure 5. Distributions of emotion valences for (A) fear, (B) sadness, (C) anger, and (D) disgust between January and October 2020.



Discussion

Principal Findings

We found that the public opinion about COVID-19 vaccines fell under 3 latent topics among Australian Twitter users from January 22 to October 20, 2020. Topic 1 was about different attitudes and actions toward COVID-19 and its vaccination. Provaxxers recognized the consequences of the COVID-19 pandemic and supported vaccine trials. Those who were skeptical about vaccines were affected by misinformation and adverse effects, which are statistically rare. Some Twitter users gave low priority to COVID-19 and hence vaccination against it and other unrelated problems. Topic 2 showed that some Twitter users advocated for infection control measures, had confidence in COVID-19 vaccine trials, and rebutted tweets that were derived from conspiracy theories or misinformation. They argued that infections and deaths from COVID-19 had overtaken previous pandemics, and other measures such as wearing masks and social distancing should be followed when mass vaccination is yet to come. Topic 3 centered on baseless claims, conspiracy theories, complaints, and misconceptions about various measures against COVID-19, including vaccines, drugs, virus testing, lockdown, and herd immunity. The major pitfall of these tweets was that their content could not be supported with any valid scientific evidence; further, the complaints were not directly associated with any solutions. Another significant finding was that nearly two-thirds of the sentiments in the tweets related to COVID-19 vaccines were positive. Of those tweets analyzed, 17% of the emotions were linked with trust and 14% were associated with anticipation. However, 14% contained fear and 8% expressed sadness. Overall, less than one-third of the tweets' sentiments were classified as negative, and one-third of the tweets were associated with the 4 negative emotions (ie, fear, sadness, anger, and disgust).

Comparison With Prior Work

In the past decade, machine learning has been applied to explore topics and sentiments of content from Twitter users about vaccinations. Some studies have examined tweets related to vaccinations in general, while others have analyzed vaccination-related tweets focusing on a particular virus or disease, such as the influenza virus, which causes respiratory illness, or HPV, which is mainly sexually transmitted. Those studies identified both positive and negative sentiments toward vaccinations, as well as neutral sentiment. Nevertheless, the outcomes of sentiment categories and the topics identified from Twitter users varied across studies focusing on different countries, years, viruses, and thus diseases.

For example, Jamison et al [85] generated 100 topics using LDA in which nearly half were annotated as provaccination, and less than 30% were coded as antivaccination from English, vaccine-relevant tweets between 2014 and 2017. However, Raghupathi et al [60] found that both positive and negative sentiments accounted for 40% of English tweets in the first half of 2019. On the other hand, the composition of sentiments in non-English tweets could be different from of English tweets. In Italy, Tavoschi et al [51] used support vector machine to classify tweets' term frequency–inverse document frequency between 2016 and 2017, and found that 60% were neutral, 23% were against vaccination, and only 17% were provaccination. It was also found that the number of provaccine tweets became greater than the number of antivaccine tweets when news about compulsory vaccination and the soaring rate of positive cases or deaths were broadcast [86].

The topics identified were not entirely similar across studies. For instance, Jamison et al [85] summarized 5 provaccine themes and 5 antivaccine themes from 100 topics; and Raghupathi et al [60] identified 3 focus areas (eg, the search for better vaccines, the disease outbreak, and debates between provaxxers and antivaxxers regarding measles). Chan et al [74],

XSL•FO
RenderX

who studied influenza vaccination in the United States, used LDA to create 10 topics in which some shared similar attributes with the themes of Jamison et al [85], such as vaccine science, safety concerns, and conspiracy theories. Some, but not all, of the similar themes, focus areas, and topics could also be seen in the analyses of tweets about vaccination regardless of virus types such as those in the studies surrounding HPV vaccinations [73,87-90].

Added Value of This Study

This study is the first topic modeling and sentiment analysis of tweets in Australia about COVID-19 vaccinations. As COVID-19 has turned into a pandemic, it is necessary to explore and summarize public opinion and sentiments pertaining to discussions on the COVID-19 vaccine, so as to prepare for the promotion of vaccination, which needs to be strengthened. This study used a traditional natural language processing technique—LDA—to identify 3 latent topics in the tweets associated with COVID-19 vaccinations: (1) attitudes toward COVID-19 and its vaccination, (2) advocating infection control measures against COVID-19, and (3) misconceptions and complaints about COVID-19 control. Furthermore, this study discovered that positive sentiment in COVID-19 vaccine discussions was higher than negative sentiment, and trust and anticipation comprised relatively large proportions of the emotions observed, as well as fear. This study visualized results using word clouds, counts of word pairs, and correlations between words, which offer supplementary angles in interpreting the results. For example, high-frequency words and word pairs that commonly appeared together were intuitively presented.

The Australian population has been the focus of research on tweets related to vaccination in previous studies. Taking the HPV vaccine as an example, nearly one-fifth of Australian Twitter users expressed health concerns about the vaccine [88], and around one-third of the exposure to information on Twitter was associated with misinformation or adverse effects of the vaccine [89]. Our study provides new insights into topics of discussion in Australia and sentiments toward vaccination against COVID-19, which is now a global pandemic and has caused over 900 deaths in Australia [3] and over 1.8 million deaths worldwide [4] as of early January 2021. By assessing public opinion and the sentiments associated with COVID-19 vaccination, governments and health agencies can plan, tailor, and implement a timely promotion of vaccination to achieve herd immunity as soon as possible.

Implications

In the results of the previous studies, we did not see a prevalent objection or opposition, in terms of topics identified or sentiments, toward vaccination regardless of virus types. A number of topics' focus areas or themes shared a certain level of similarity across studies concerning different viruses. For instance, topics of safety, scientific evidence, and conspiracy theories were commonly found across studies. Topics like scandals associated with vaccines, misinformation, and disease outbreaks were identified in some other studies. These results indicated public concern about the benefits and risks of vaccination at the individual and social levels, and the type of virus or disease when deciding whether to receive a vaccine or not.

In our study, besides fabricated information such as microchips in vaccines and the flu vaccine causing COVID-19 deaths, some Twitter users thought that COVID-19 was not serious enough compared to other existing global crises, and that the pandemic was being politicized or commercialized. These conspiracy theories, along with other antivaccine propagandas such as encouraging natural herd immunity, indicated that the risks of deaths, complications, or sequela arising from COVID-19 to others, or to oneself, were acceptable to some members of the public.

Although the Australian opinion showed more positive sentiment related to COVID-19 vaccinations, the positive sentiment was not a leading majority compared to the negative one. This means more work needs to be done to promote vaccination so as to achieve herd immunity to protect vulnerable and minority groups. Rigorous science that is easily understandable needs to replace biased, fabricated, or outdated information in the public. Governments should build and strengthen the public's confidence in COVID-19 vaccination, if it is not mandatory, that is, required by law, beyond arranging vaccine delivery logistically and vaccine administration clinically.

Limitations

Our results represent Twitter users in the Australian public, which is a different approach from national survey statistics. However, the public opinions collected on Twitter may represent views from younger populations. Previous studies showed that around 85% to 90% of Twitter users were aged less than 25-40 years, which varied across locations such as the Netherlands [91], the United Kingdom [92] and other places [93]. Older adults' opinions require further investigations with modifications to the study design whereas younger adults' opinions on the vaccine deserve continuous attention. Goldstein et al [94] reported that those aged less than 35 years had high cumulative rates of COVID-19 infections in the community where transmissions in secondary schools or high schools were robust. A report published by the US Centers for Disease Control and Prevention [95] showed that the percent positivity of SARS-CoV-2 RT-PCR (reverse transcription–polymerase chain reaction) tests increased early among young people, followed by a rise in positivity in middle-aged and older adults. Consequently, around 20% of adolescents manifested symptoms compared with nearly 70% of the elderly [96], who are subject to a higher probability of further developments leading to death. Hence, there is an urgent need to explore younger population's opinion and acceptability of vaccination, which could have significant impact on disease control in the first place.

In addition to the study period and the country of concern, analysis methods might lead to variation in topics and sentiments toward vaccinations. For supervised learning such as support vector machine, a training set is required, which needs to be manually labeled; this might carry some subjectivity in categorizing tweets into predefined topics for training. However, the advantage is that the set could be used to validate the model performance and then test a large data set. Considering unsupervised learning such as LDA, Dirichlet multinomial

XSL•FO
RenderX

mixtures (DMM), and k-means of term frequency–inverse document frequency, the primary limitation is the subjectivity in defining the topics created [60,74]. In addition, a sound reason or calculation is needed to support the preset number of topics, which would affect the results.

Some previous studies generated a rather high number of topics (30-100) using an LDA or DMM model, and then manually grouped the topics into themes [73,85,89]. However, there was risk of bias since the content of each topic was not reported in detail, and the contents of the themes could be mixed, which is difficult to interpret. Furthermore, the manual grouping also contained the risk of subjectivity. In the current study, we adopted LDA, which was similar to the one used by Chan et al [74]. We identified 3 latent topics in which the importance of words were visualized; the frequency of word pairs and correlations between words provided additional results corresponding to the topic content.

Regarding sentiment analysis, the number of emotion categories were limited to 8 [83,97], but emotion is an abstract and broad concept that may involve as many as 27 categories [98]. Furthermore, words with spelling mistakes could not be identified and analyzed in the algorithm. With respect to each term for the development of an emotion lexicon by Mohammad and Turney [83], only 5 individuals in the public were recruited to annotate a term against each of the 8 emotions. The emotions of a term were annotated without considering possible contexts. Moreover, the interrater reliability statistics were not reported though the agreement percentages were apparently high.

Future Directions

Our study adopted an unsupervised machine learning method—LDA—for topic analysis. Future studies could investigate supervised learning to train classifiers to categorize tweets into different topics and sentiments based on a recognized theoretical framework. Such a framework could be proposed after an extensive literature review and qualitative synthesis; manual annotations should be as transparent, objective, and reliable as possible. Results from supervised learning following the same theoretical framework could be compared across the analyses of different data sets, for example, the results from different countries as shown by Shapiro et al [88]. Public opinions across countries require further study. For instance, recent online surveys of US adults found that only half claimed that they were “very likely” to get the COVID-19 vaccine [99], and one-third would not accept recommendations for vaccination [100]. In the United Kingdom, around one-third of the adult sample showed hesitancy or resistance against COVID-19 vaccination [101,102]. In the future, a spatiotemporal analysis of tweets about COVID-19 vaccination could be attempted. Similar studies have been conducted on Twitter data to study emergency department visits for influenza-like illness in New York City [103], COVID-19–related stress symptoms in the United States [104], and communicating the risk of MERS infections in South Korea [105]. Furthermore, individual reactions toward the COVID-19 vaccine in tweets could be monitored over time and tested for correlations between frequencies of identified topics or emotions, important real events, and health indicators such as vaccination coverage, infection rate, and death rate. In addition to studying the spread of misinformation and conspiracy theories on social media, future research should explore personal values that might hinder collective health care strategies and positive outcomes.

Conclusions

Our findings indicate that the Australian public possessed varying attitudes toward COVID-19 and its vaccination. Moreover, some had misconceptions and complaints about COVID-19 and infection control measures, while others advocated for pharmaceutical and nonpharmacological measures against COVID-19. Nonetheless, in our sentiment analysis, the level of positive sentiment in public opinion may not be strong enough to further a high vaccination coverage to achieve vaccination-induced herd immunity, which is essential to protect oneself and others. For those without contraindications, getting vaccinated is not merely a personal choice but is also a way of protecting the community. Governments should explore public opinion and sentiments toward COVID-19 vaccination and get the public psychologically prepared for vaccination with evidence-based, authorized, and understandable information, in addition to supporting the biomedical development, storage, delivery, and clinical administration of vaccines.

Acknowledgments

This work was supported in part by the Murdoch University New Staff Startup Grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of tweets collected between January 22 and October 20, 2020.
[DOC File , 31 KB - jmir_v23i5e26953_app1.doc]

Multimedia Appendix 2
Plot of word tokens against counts sorted.
[DOC File , 24 KB - jmir_v23i5e26953_app2.doc]

XSL•FO
RenderX

Multimedia Appendix 3
Plot of word pairs against counts sorted.
[DOC File , 25 KB - jmir_v23i5e26953_app3.doc]

Multimedia Appendix 4
Plot of number of topics against LDA tuning scores.
[DOC File , 146 KB - jmir_v23i5e26953_app4.doc]

Multimedia Appendix 5
Word tokens with counts above 250 in the corpus.
[PNG File , 397 KB - jmir_v23i5e26953_app5.png]

Multimedia Appendix 6
Word distributions over 3 topics in the latent Dirichlet allocation model.
[PNG File , 439 KB - jmir_v23i5e26953_app6.png]

Multimedia Appendix 7
Top 100 probability (beta) distributions of words in each topic.
[DOCX File , 24 KB - jmir_v23i5e26953_app7.docx]

Multimedia Appendix 8
Top 20 probability (theta) distributions of topic 1 in tweet samples.
[DOCX File , 18 KB - jmir_v23i5e26953_app8.docx]

Multimedia Appendix 9
Top 20 probability (theta) distributions of topic 2 in tweet samples.
[DOCX File , 17 KB - jmir_v23i5e26953_app9.docx]

Multimedia Appendix 10
Top 20 probability (theta) distributions of topic 3 in tweet samples.
[DOCX File , 18 KB - jmir_v23i5e26953_app10.docx]

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
2. Countries/areas with reported cases of Coronavirus Disease 2019 (COVID 19) (Last updated on January 4, 2021, 11 am). Hong Kong Centre for Health Protection. 2020. URL: https://www.chp.gov.hk/files/pdf/statistics_of_the_cases_novel_coronavirus_infection_en_20210104.pdf [accessed 2021-05-06]
3. Coronavirus (COVID-19) current situation and case numbers. Australian Department of Health. 2020. URL: https://tinyurl.com/uu5semny [accessed 2021-05-06]
4. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-09-22]
5. Debata B, Patnaik P, Mishra A. COVID ‐ 19 pandemic! It's impact on people, economy, and environment. J Public Affairs 2020 Sep 02;20:e2372. [doi: 10.1002/pa.2372]
6. Song L, Zhou Y. The COVID-19 pandemic and its impact on the global economy: What does it take to turn crisis into opportunity? China & World Economy 2020 Aug 04;28(4):1-25. [doi: 10.1111/cwe.12349]
7. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. Official Journal of the Association of Medical Microbiology and Infectious Disease Canada 2020 Dec;5(4):223-234. [doi: 10.3138/jammi-2020-0030]
8. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. J Med Virol 2021 Feb 13;93(2):820-830 [FREE Full text] [doi: 10.1002/jmv.26326] [Medline: 32691881]
9. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
10. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies

XSL•FO
RenderX

from 9 countries. PLoS One 2020;15(6):e0234765 [FREE Full text] [doi: 10.1371/journal.pone.0234765] [Medline: 32574165]
11. Baksh M, Ravat V, Zaidi A, Patel RS. A systematic review of cases of acute respiratory distress syndrome in the coronavirus disease 2019 pandemic. Cureus 2020 May 18;12(5):e8188 [FREE Full text] [doi: 10.7759/cureus.8188] [Medline: 32566429]
12. Hasan SS, Capstick T, Ahmed R, Kow CS, Mazhar F, Merchant HA, et al. Mortality in COVID-19 patients with acute respiratory distress syndrome and corticosteroids use: A systematic review and meta-analysis. Expert Rev Respir Med 2020 Nov;14(11):1149-1163 [FREE Full text] [doi: 10.1080/17476348.2020.1804365] [Medline: 32734777]
13. Zhang Z, Zhang L, Wang Y. COVID-19 indirect contact transmission through the oral mucosa must not be ignored. J Oral Pathol Med 2020 May;49(5):450-451 [FREE Full text] [doi: 10.1111/jop.13019] [Medline: 32281674]
14. Alimohamadi Y, Taghdir M, Sepandi M. Estimate of the basic reproduction number for COVID-19: A systematic review and meta-analysis. J Prev Med Public Health 2020 May;53(3):151-157. [doi: 10.3961/jpmph.20.076]
15. Billah A, Miah M, Khan N. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. PLoS One 2020;15(11):e0242128 [FREE Full text] [doi: 10.1371/journal.pone.0242128] [Medline: 33175914]
16. Lei H, Xu X, Xiao S, Wu X, Shu Y. Household transmission of COVID-19-a systematic review and meta-analysis. J Infect 2020 Dec;81(6):979-997 [FREE Full text] [doi: 10.1016/j.jinf.2020.08.033] [Medline: 32858069]
17. Rehman M, Tauseef I, Aalia B, Shah SH, Junaid M, Haleem KS. Therapeutic and vaccine strategies against SARS-CoV-2: Past, present and future. Future Virology 2020 Jul;15(7):471-482. [doi: 10.2217/fvl-2020-0137]
18. Badgujar KC, Badgujar VC, Badgujar SB. Vaccine development against coronavirus (2003 to present): An overview, recent advances, current scenario, opportunities and challenges. Diabetes Metab Syndr 2020;14(5):1361-1376 [FREE Full text] [doi: 10.1016/j.dsx.2020.07.022] [Medline: 32755836]
19. Spencer JP, Trondsen Pawlowski RH, Thomas S. Vaccine adverse events: Separating myth from reality. American Academy of Family Physicians 2017;95(12):786-794 [FREE Full text]
20. Maglione MA, Das L, Raaen L, Smith A, Chari R, Newberry S, et al. Safety of vaccines used for routine immunization of U.S. children: A systematic review. Pediatrics 2014 Aug;134(2):325-337. [doi: 10.1542/peds.2014-1079] [Medline: 25086160]
21. Fine P, Eames K, Heymann DL. "Herd immunity": A rough guide. Clin Infect Dis 2011 Apr 1;52(7):911-916 [FREE Full text] [doi: 10.1093/cid/cir007] [Medline: 21427399]
22. Lopez A, Mariette X, Bachelez H, Belot A, Bonnotte B, Hachulla E, et al. Vaccination recommendations for the adult immunosuppressed patient: A systematic review and comprehensive field synopsis. J Autoimmun 2017 Jun;80:10-27. [doi: 10.1016/j.jaut.2017.03.011] [Medline: 28381345]
23. Bansal P, Goyal A. COVID-19 – Challenges ahead of vaccination in immunocompromised patients. Gen Int Med Clin Innov 2020;5:1-3. [doi: 10.15761/gimci.1000196]
24. Hussain A, Ali S, Ahmed M, Hussain S. The anti-vaccination movement: A regression in modern medicine. Cureus 2018 Jul 03;10(7):e2919 [FREE Full text] [doi: 10.7759/cureus.2919] [Medline: 30186724]
25. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998 Feb 28;351(9103):637-641. [Medline: 9500320]
26. Taylor B, Miller E, Farrington CP, Petropoulos MC, Favot-Mayaud I, Li J, et al. Autism and measles, mumps, and rubella vaccine: No epidemiological evidence for a causal association. Lancet 1999 Jun 12;353(9169):2026-2029. [doi: 10.1016/s0140-6736(99)01239-8] [Medline: 10376617]
27. Fombonne E, Chakrabarti S. No evidence for a new variant of measles-mumps-rubella-induced autism. Pediatrics 2001 Oct 01;108(4):e58. [doi: 10.1542/peds.108.4.e58]
28. Farrington CP, Miller E, Taylor B. MMR and autism: Further evidence against a causal association. Vaccine 2001 Jun 14;19(27):3632-3635. [doi: 10.1016/s0264-410x(01)00097-4] [Medline: 11395196]
29. DeStefano F, Thompson WW. MMR vaccine and autism: An update of the scientific evidence. Expert Rev Vaccines 2004 Feb;3(1):19-22. [doi: 10.1586/14760584.3.1.19] [Medline: 14761240]
30. Peltola H, Patja A, Leinikki P, Valle M, Davidkin I, Paunio M. No evidence for measles, mumps, and rubella vaccine-associated inflammatory bowel disease or autism in a 14-year prospective study. The Lancet 1998 May;351(9112):1327-1328. [doi: 10.1016/s0140-6736(98)24018-9]
31. DeStefano F, Chen RT. Negative association between MMR and autism. The Lancet 1999 Jun;353(9169):1987-1988. [doi: 10.1016/s0140-6736(99)00160-9]
32. The Editors of The Lancet. Retraction—Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. The Lancet 2010 Feb;375(9713):445. [doi: 10.1016/s0140-6736(10)60175-4]
33. Singh K, Mehta S. The clinical development process for a novel preventive vaccine: An overview. J Postgrad Med 2016;62(1):4-11 [FREE Full text] [doi: 10.4103/0022-3859.173187] [Medline: 26732191]
34. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, Oxford COVID Vaccine Trial Group. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet 2020 Aug 15;396(10249):467-478 [FREE Full text] [doi: 10.1016/S0140-6736(20)31604-4] [Medline: 32702298]

XSL•FO
RenderX

35. Zhu F, Guan X, Li Y, Huang J, Jiang T, Hou L, et al. Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: A randomised, double-blind, placebo-controlled, phase 2 trial. The Lancet 2020 Aug;396(10249):479-488. [doi: 10.1016/s0140-6736(20)31605-6]
36. The push for a COVID-19 vaccine. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines [accessed 2020-09-22]
37. Immunization schedules. US Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/vaccines/schedules/hcp/schedule-changes.html [accessed 2021-05-06]
38. FDA insight: Vaccines for COVID-19, Part 1. US Food and Drug Administration. 2020. URL: https://www.fda.gov/news-events/fda-insight/fda-insight-vaccines-covid-19-part-1 [accessed 2021-05-06]
39. Distributing vaccines and treatments for COVID-19 and flu. UK Department of Health and Social Care. 2020. URL: https://www.gov.uk/government/consultations/distributing-vaccines-and-treatments-for-covid-19-and-flu [accessed 2021-05-06]
40. Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. J Med Internet Res 2008 Jun 30;10(2):e17 [FREE Full text] [doi: 10.2196/jmir.986] [Medline: 18653441]
41. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
42. Smallpox vaccines. World Health Organization. 2016 May 31. URL: https://www.who.int/csr/disease/smallpox/vaccines/en/ [accessed 2021-05-06]
43. Khuroo MS, Khuroo M, Khuroo MS, 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]
44. Funk CD, Laferrière C, Ardakani A. A snapshot of the global race for vaccines targeting SARS-CoV-2 and the COVID-19 pandemic. Front Pharmacol 2020;11:937 [FREE Full text] [doi: 10.3389/fphar.2020.00937] [Medline: 32636754]
45. Dodd RH, Cvejic E, Bonner C, Pickles K, McCaffery KJ, Ayre J, et al. Willingness to vaccinate against COVID-19 in Australia. The Lancet Infectious Diseases 2020 Jun;21:318-319. [doi: 10.1016/s1473-3099(20)30559-4]
46. New deal secures potential COVID-19 vaccine for every Australian. Prime Minister of Australia. Canberra: Australian Government; 2020 Aug 21. URL: https://www.health.gov.au/news/new-deal-secures-potential-covid-19-vaccine-for-every-australian-0 [accessed 2021-05-06]
47. Cyranoski D, Mallapaty S. Scientists relieved as coronavirus vaccine trial restarts - but question lack of transparency. Nature 2020 Sep 14;585(7825):331-332. [doi: 10.1038/d41586-020-02633-6] [Medline: 32929259]
48. Mahase E. Covid-19: Johnson and Johnson vaccine trial is paused because of unexplained illness in participant. BMJ 2020 Oct 13;371:m3967. [doi: 10.1136/bmj.m3967] [Medline: 33051241]
49. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
50. Deiner MS, Fathy C, Kim J, Niemeyer K, Ramirez D, Ackley SF, et al. Facebook and Twitter vaccine sentiment in response to measles outbreaks. Health Informatics J 2019;25(3):1116-1132. [doi: 10.1177/1460458217740723] [Medline: 29148313]
51. Tavoschi L, Quattrone F, D'Andrea E, Ducange P, Vabanesi M, Marcelloni F, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from September 2016 to August 2017 in Italy. Human Vaccines & Immunotherapeutics 2020 Mar 02;16(5):1062-1069. [doi: 10.1080/21645515.2020.1714311]
52. Ortiz-Sánchez E, Velando-Soriano A, Pradas-Hernández L, Vargas-Román K, Gómez-Urquiza JL, Cañadas-De la Fuente GA, et al. Analysis of the Anti-Vaccine Movement in Social Networks: A Systematic Review. Int J Environ Res Public Health 2020 Jul 27;17(15):5394 [FREE Full text] [doi: 10.3390/ijerph17155394] [Medline: 32727024]
53. Megget K. Even covid-19 can't kill the anti-vaccination movement. BMJ 2020 Jun 04;369:m2184. [doi: 10.1136/bmj.m2184] [Medline: 32499217]
54. Bruns A, Harrington S, Hurcombe E. 'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook. Media International Australia 2020 Aug 04;177(1):12-29. [doi: 10.1177/1329878x20946113]
55. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
56. Investor fact sheet. Twitter. 2020. URL: https://s22.q4cdn.com/826641620/files/doc_financials/2019/q4/Q4_19_InvestorFactSheet.pdf [accessed 2021-05-06]
57. Khanna P. An approach to opinion mining from Twitter data using R. IJARCS 2017 Aug 30;8(8):252-256. [doi: 10.26483/ijarcs.v8i8.4716]
58. Zimbra D, Abbasi A, Zeng D, Chen H. The State-of-the-Art in Twitter Sentiment Analysis. ACM Trans Manage Inf Syst 2018 Sep 05;9(2):1-29. [doi: 10.1145/3185045]
59. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 03;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]

XSL•FO
RenderX

60. Raghupathi V, Ren J, Raghupathi W. Studying public perception about vaccination: A sentiment analysis of tweets. IJERPH 2020 May 15;17(10):3464. [doi: 10.3390/ijerph17103464]
61. Kearney M. rtweet: Collecting and analyzing Twitter data. JOSS 2019 Oct;4(42):1829. [doi: 10.21105/joss.01829]
62. Rinker TW. qdapRegex: Regular expression removal, extraction, and replacement tools, version 0.7.2. qdapRegex. Buffalo, NY: University at Buffalo; 2017. URL: http://github.com/trinker/qdapRegex [accessed 2021-05-06]
63. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Stat Soft 2008;25(5):1-54. [doi: 10.18637/jss.v025.i05]
64. Feinerer I, Hornik K. tm: Text mining package. R package, version 0.7-8. CRAN. 2020. URL: https://CRAN.R-project.org/package=tm [accessed 2021-05-06]
65. Honnibal M, Montani I, Van Landeghem S, Boyd A. spaCy: Industrial-strength natural language processing in Python. Zenodo. 2020. URL: https://doi.org/10.5281/zenodo.1212303 [accessed 2020-12-23]
66. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. : ELRA; 2010 Presented at: LREC Workshop on New Challenges for NLP Frameworks; May 22; Valletta, Malta p. 45-50.
67. Sedgewick R, Wayne K. 3.5 Searching Applications. In: Algorithms, 4th edition. New Jersey: Addison-Wesley Professional; 2014.
68. Abstract service stopword list. SAO/NASA Astrophysics Data System. 2020. URL: http://adsabs.harvard.edu/abs_doc/stopwords.html [accessed 2020-11-07]
69. Robinson D, Misra K, Silge J. widyr: Widen, process, and re-tidy a dataset. CRAN. 2020. URL: https://cran.r-project.org/web/packages/widyr/index.html [accessed 2021-05-06]
70. Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
71. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
72. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research 2003;3(4-5):993-1022.
73. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
74. Chan MS, Jamieson KH, Albarracin D. Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States. Vaccine 2020 Sep 11;38(40):6236-6247 [FREE Full text] [doi: 10.1016/j.vaccine.2020.07.054] [Medline: 32792251]
75. Nikita M, Chaney N. ldatuning: Tuning of the latent Dirichlet allocation models parameters. CRAN. 2020. URL: https://cran.r-project.org/web/packages/ldatuning/index.html [accessed 2021-05-06]
76. Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN. On finding the natural number of topics with latent Dirichlet allocation: Some observations. In: Zaki MJ, Yu JX, Ravindran B, Pudi V, editors. Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science, vol 6118. Berlin, Heidelberg: Springer; 2010:391-402.
77. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7-9):1775-1781. [doi: 10.1016/j.neucom.2008.06.011]
78. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci USA 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
79. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document Numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]
80. Grün B, Hornik K. Topicmodels: An R package for fitting topic models. J Stat Soft 2011;40(13):1-30. [doi: 10.18637/jss.v040.i13]
81. Jockers ML. Syuzhet: Extract sentiment and plot arcs from text. GitHub. 2015. URL: https://github.com/mjockers/syuzhet [accessed 2021-05-06]
82. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP natural language processing toolkit. 2014 Presented at: 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; June; Baltimore, Maryland p. 55-60. [doi: 10.3115/v1/p14-5010]
83. Mohammad S, Turney P. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. 2010 Presented at: NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text; June; Los Angeles, CA.
84. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Computational Intelligence 2013;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]
85. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and extending a typology to identify vaccine misinformation on Twitter. Am J Public Health 2020 Oct;110(S3):S331-S339. [doi: 10.2105/ajph.2020.305940]
86. D'Andrea E, Ducange P, Bechini A, Renda A, Marcelloni F. Monitoring the public opinion about the vaccination topic from tweets analysis. Expert Systems with Applications 2019 Feb;116:209-226. [doi: 10.1016/j.eswa.2018.09.009]

XSL•FO
RenderX

87. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
88. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017 Oct 05;7(10):e016869 [FREE Full text] [doi: 10.1136/bmjopen-2017-016869] [Medline: 28982821]
89. Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, et al. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Hum Vaccin Immunother 2019;15(7-8):1488-1495 [FREE Full text] [doi: 10.1080/21645515.2019.1596712] [Medline: 30978147]
90. Luo X, Zimet G, Shah S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in Twitter over 10 years (2008 - 2017). Human Vaccines & Immunotherapeutics 2019 Jul 16;15(7-8):1496-1504. [doi: 10.1080/21645515.2019.1627821]
91. Nguyen D, Gravel R, Trieschnigg RB, Meder T. "How old do you think I am?": A study of language and age in Twitter. Palo Alto, California: AAAI Press; 2013:439-448.
92. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PLoS One 2015;10(3):e0115545 [FREE Full text] [doi: 10.1371/journal.pone.0115545] [Medline: 25729900]
93. Morgan-Lopez AA, Kim AE, Chew RF, Ruddle P. Predicting age groups of Twitter users based on language and metadata features. PLoS One 2017;12(8):e0183537 [FREE Full text] [doi: 10.1371/journal.pone.0183537] [Medline: 28850620]
94. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. J Infect Dis 2020;223(3):362-369 [FREE Full text] [doi: 10.1093/infdis/jiaa691] [Medline: 33119738]
95. Oster AM, Caruso E, DeVies J, Hartnett KP, Boehmer TK. Transmission dynamics by age group in COVID-19 hotspot counties - United States, April-September 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 16;69(41):1494-1496 [FREE Full text] [doi: 10.15585/mmwr.mm6941e1] [Medline: 33056949]
96. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med 2020 Aug;26(8):1205-1211. [doi: 10.1038/s41591-020-0962-9] [Medline: 32546824]
97. Plutchik R. A general psychoevolutionary theory of emotion. Emotion: Theory, Research, and Experience 1980;1(3):3-33. [doi: 10.1016/b978-0-12-558701-3.50007-7]
98. Cowen AS, Keltner D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc Natl Acad Sci USA 2017 Sep 19;114(38):E7900-E7909 [FREE Full text] [doi: 10.1073/pnas.1702247114] [Medline: 28874542]
99. Khubchandani J, Sharma S, Price JH, Wiblishauser MJ, Sharma M, Webb FJ. COVID-19 vaccination hesitancy in the United States: A rapid national assessment. J Community Health 2021 Apr;46(2):270-277 [FREE Full text] [doi: 10.1007/s10900-020-00958-x] [Medline: 33389421]
100. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
101. Paul E, Steptoe A, Fancourt D. Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications. The Lancet Regional Health - Europe 2021 Feb;1:100012. [doi: 10.1016/j.lanepe.2020.100012]
102. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nat Commun 2021 Jan 04;12(1):29. [doi: 10.1038/s41467-020-20226-9] [Medline: 33397962]
103. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014 Oct 20;16(10):e236-e274 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]
104. Li D, Chaudhary H, Zhang Z. Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining. IJERPH 2020 Jul 10;17(14):4988. [doi: 10.3390/ijerph17144988]
105. Kim I, Feng C, Wang Y, Spitzberg BH, Tsou M. Exploratory spatiotemporal analysis in risk communication during the MERS outbreak in South Korea. The Professional Geographer 2017 Mar 31;69(4):629-643. [doi: 10.1080/00330124.2017.1288577]

Abbreviations

API: application programming interface
DMM: Dirichlet multinomial mixtures
HPV: human papillomavirus
LDA: latent Dirichlet allocation
MERS: Middle East respiratory syndrome
MERS-CoV: Middle East respiratory syndrome coronavirus
MMR: measles, mumps, and rubella

XSL•FO
RenderX

NASA: National Aeronautics and Space Administration
R_0: reproductive number
NLP: natural language processing
PCV13: 13-valent pneumococcal conjugate vaccines
RT-PCR: reverse transcription–polymerase chain reaction
SAO: Smithsonian Astrophysical Observatory
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus

Edited by C Basch; submitted 05.01.21; peer-reviewed by P Qian, Y Jiang; comments to author 27.01.21; revised version received 02.02.21; accepted 16.04.21; published 19.05.21.

Please cite as:
Kwok SWH, Vadde SK, Wang G
Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis
J Med Internet Res 2021;23(5):e26953
URL: https://www.jmir.org/2021/5/e26953
doi:10.2196/26953
PMID:33886492

©Stephen Wai Hang Kwok, Sai Kumar Vadde, Guanjin Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Characterizing Health Care Delays and Interruptions in the United States During the COVID-19 Pandemic: Internet-Based, Cross-sectional Survey Study

Elizabeth Lerner Papautsky[1], BSc, MSc, PhD; Dylan R Rice[2], BA; Hana Ghoneima[3]; Anna Laura W McKowen[3], MA; Nicholas Anderson[3], BA, MSc; Angie R Wootton[4], MSW; Cindy Veldhuis[3], PhD

[1]Department of Biomedical & Health Information Sciences, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

[2]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

[3]Department of Counseling & Clinical Psychology, Teachers College, Columbia University, New York, NY, United States

[4]School of Social Welfare, University of California Berkeley, Berkeley, CA, United States

Corresponding Author:
Elizabeth Lerner Papautsky, BSc, MSc, PhD
Department of Biomedical & Health Information Sciences
College of Applied Health Sciences
University of Illinois at Chicago
1919 W Taylor St
Chicago, IL, 60612
United States
Phone: 1 312 996 0357
Email: elp@uic.edu

Abstract

Background: The COVID-19 pandemic has broader geographic spread and potentially longer lasting effects than those of previous disasters. Necessary preventive precautions for the transmission of COVID-19 has resulted in delays for in-person health care services, especially at the outset of the pandemic.

Objective: Among a US sample, we examined the rates of delays (defined as cancellations and postponements) in health care at the outset of the pandemic and characterized the reasons for such delays.

Methods: As part of an internet-based survey that was distributed on social media in April 2020, we asked a US–based convenience sample of 2570 participants about delays in their health care resulting from the COVID-19 pandemic. Participant demographics and self-reported worries about general health and the COVID-19 pandemic were explored as potent determinants of health care delays. In addition to all delays, we focused on the following three main types of delays, which were the primary outcomes in this study: dental, preventive, and diagnostic care delays. For each outcome, we used bivariate statistical tests (*t* tests and chi-square tests) and multiple logistic regression models to determine which factors were associated with health care delays.

Results: The top reported barrier to receiving health care was the fear of SARS-CoV-2 infection (126/374, 33.6%). Almost half (1227/2570, 47.7%) of the participants reported experiencing health care delays. Among those who experienced health care delays and further clarified the type of delay they experienced (921/1227, 75.1%), the top three reported types of care that were affected by delays included dental (351/921, 38.1%), preventive (269/921, 29.2%), and diagnostic (151/921, 16.4%) care. The logistic regression models showed that age (P<.001), gender identity (P<.001), education (P=.007), and self-reported worry about general health (P<.001) were significantly associated with experiencing health care delays. Self-reported worry about general health was negatively related to experiencing delays in dental care. However, this predictor was positively associated with delays in diagnostic testing based on the logistic regression model. Additionally, age was positively associated with delays in diagnostic testing. No factors remained significant in the multiple logistic regression for delays in preventive care, and although there was trend between race and delays (people of color experienced fewer delays than White participants), it was not significant (P=.06).

Conclusions: The lessons learned from the initial surge of COVID-19 cases can inform systemic mitigation strategies for potential future disruptions. This study addresses the demand side of health care delays by exploring the determinants of such delays. More research on health care delays during the pandemic is needed, including research on their short- and long-term

XSL•FO
RenderX

impacts on patient-level outcomes such as mortality, morbidity, mental health, people's quality of life, and the experience of pain.

(J Med Internet Res 2021;23(5):e25446) doi:10.2196/25446

KEYWORDS

COVID-19; health care delays; internet survey; preventive care; delay; interruption; lockdown; precaution; prevention; social media; survey

Introduction

Times of disaster have often been characterized by a lack of access to health care, treatment delays or interruptions, and medication shortages [1-4]. The COVID-19 pandemic has broader geographic spread and potentially longer lasting effects than those of previous disasters. Further, the care of patients with COVID-19 has overburdened health care systems, and seeking health care may put patients at risk of exposure to SARS-CoV-2. Implementing necessary infection prevention precautions for the transmission of SARS-CoV-2 (the virus that causes COVID-19) and allocating resources (eg, personal protective equipment, beds, and personnel) in anticipation of or in response to a surge of COVID-19 cases has resulted in delays in and cancellations of in-person health care services, especially at the outset of the pandemic. These delays may affect a large proportion of patients. For example, in a recent study about health care delays at the outset of the pandemic among breast cancer survivors (April 2 to April 27, 2020; N=609), 44% of participants reported breast cancer treatment delays [5]. The consequences can be significant for individuals with conditions that require timely intervention, such as cancer, other chronic health conditions, and mental health concerns. The deleterious effects of postponing preventive, diagnostic, and dental health care may include the delayed or missed diagnosis of life-threatening illnesses such as cancer, the exacerbation of illnesses, and even death. Given the potential implications for the morbidity and mortality of delaying or not accessing health care, it is important to document delays and interruptions in health care access during the COVID-19 pandemic to inform appropriate intervention efforts.

Of great interest and concern are delays in preventive and diagnostic care, particularly delays in such care for older adults. Before the pandemic, approximately 50% of all adults received the recommended preventive care in the United States [6-8]. Thus, people with pre-existing and emerging conditions may not be receiving needed health care management—a problem that was not introduced but rather amplified by the pandemic. Further, as preventive and diagnostic care are the first line of defense for diagnosing life-threatening illness, delays and interruptions may result in higher mortality and morbidity. This is of particular concern for older adults, given that age is the main risk factor for cardiovascular disease, cancer, and neurodegenerative conditions [9], and such conditions may be exacerbated by stress and isolation during the pandemic. Emerging studies on delayed and interrupted care have focused on specialties such as head and neck malignancies [10], urologic surgeries [11], and heart failure [12]. Ding and colleagues [13] documented the potentially fatal consequences of care delays for COVID-19–negative patients requiring in-person medical care for timely diagnosis. They described five pediatric patient cases with poor clinical outcomes resulting from delays in cancer diagnosis. These poor outcomes unfortunately occurred despite the fact that several patients interacted with primary care providers through telehealth visits prior to presenting at the hospital. Due to the effects of the COVID-19 pandemic on people's access to health care, the oncology community anticipates an uptick in the number of cancer diagnoses and diagnoses at higher cancer stages and an increase in cancer mortality rates overall [14]—a devastating and sobering concern. Research with a focus on older adults is needed to characterize the near- and long-term impacts of delayed and interrupted care on health and inform effective and person-centered solutions to protect severely ill but treatable patients.

To address a gap in our understanding of the effects of the pandemic on health care access, we examined the rates of delays in health care among a US convenience sample at the outset of the pandemic, with a specific focus on preventive care. Our study reflects Americans' experiences with the demand for health care. We also characterized the reasons for such delays. As we navigate the pandemic and the potential resurgences of infections, our findings can inform systems-level strategies for future public health emergencies to mitigate delays, improve effective communications for addressing uncertainty, and promote mental health interventions.

Methods

Study Design

In April 5, 2020, we launched an internet-based survey study examining the impact of the COVID-19 pandemic on mental health and well-being. Participants were invited to take part in a study that focused on "how the pandemic is affecting you" through social media (ie, Twitter and Facebook), listservs, social networks, websites (eg, Buzzfeed), and the research match website at Columbia University. Information about this study was included in the beginning of the survey; participants indicated consent by proceeding to the survey. The survey comprised multiple validated instruments as well as items that were developed to help us understand experiences that were specific to the COVID-19 pandemic. The mean completion time for the full survey was approximately 35 minutes. As this study was not funded, participants were not offered compensation. At the culmination of the survey, participants were provided with a list of resources (eg, crisis hotlines) in case they experienced distress. Participants were also asked to provide their email addresses if they consented to being recontacted for future follow-ups. Participants who provided their email addresses were surveyed again between April 26 and May 5,

XSL•FO
RenderX

2020. All study procedures were reviewed and approved by the Columbia University Institutional Review Board.

Measures

Participants were asked about any interruptions they experienced to their health care in the following item, which was developed specifically for this study: "Have you experienced any delays or interruptions in your healthcare (e.g. cancelled or delayed appointments, tests, procedures) during the coronavirus outbreak?" Response options for this item were dichotomous ("yes" or "no"), and participants who responded affirmatively were asked to elaborate with an open-ended, text-based answer. Responses ranged from no delayed care to multiple types of delayed care. Some participants included more detailed responses.

Data Coding and Analysis

To characterize and code the types of delays, we used Microsoft Excel. Three researchers (ELP, DR, and HG) coded participants' responses to the open-ended item about the types of health care delays and interruptions they experienced [15,16]. The coding process was as follows. First, one researcher (DR) reviewed a set of 25 randomly selected responses and developed a draft codebook, which primarily focused on the types of care affected. Second, the three researchers collaboratively reviewed the codebook draft and planned the logistics of the coding process. Each researcher coded the same 50 responses, identified coding uncertainties, and noted potential codebook edits. To establish interrater reliability, the three researchers discussed coding discrepancies until they were resolved with 100% agreement among all three researchers. The codebook was edited to clarify codes, add additional codes, and remove codes. Third, each researcher coded a third of the remaining responses. Any uncertainties were resolved by communication and discussion with the other two researchers. Finally, upon completion, researchers debriefed to further refine the codebook (eg, adding a dermatology code based on its frequency of appearance in the data set). All codes (n=2776) were integrated into a single spreadsheet.

Statistical Analysis

Basic descriptive statistics were used to describe the research sample. We conducted two sets of inferential procedures to determine which factors were associated with self-reported health care delays, with a focus on three outcome variables. First, we examined the determinants of experiencing any health care delays due to the pandemic (dichotomous variable; "yes" or "no"). Among those who experienced delays, we were particularly interested in the following three types of delays: dental, preventive, and diagnostic care delays. For each dependent variable, bivariate analyses were conducted to describe the associations of the outcomes with demographic characteristics or self-reported levels of worry about their general health and COVID-19. Independent samples *t* tests were used for continuous independent variables, whereas chi-square tests were used for categorical variables. All independent variables were then entered into a multiple logistic regression model; R (The R Foundation) [17] was used for the statistical inference procedures.

Results

The analytic sample for this study included 2570 participants (accounting for missing data in the delays question) from across all US states who completed the survey between April 5, 2020, and May 5, 2020. Participants' ages ranged from 18 to 84 years (mean 37.3 years, SD 12.6 years). The majority of participants were non-Hispanic White (2464/2570, 95.9%), were cisgender (ie, gender identity is consistent with the sex assigned at birth) women (2456/2570, 95.6%), were heterosexual (1680/2570, 65.4%), and had at least a Bachelor's degree (1259/2570, 50%). Table 1 shows a summary of demographics.

XSL•FO
RenderX

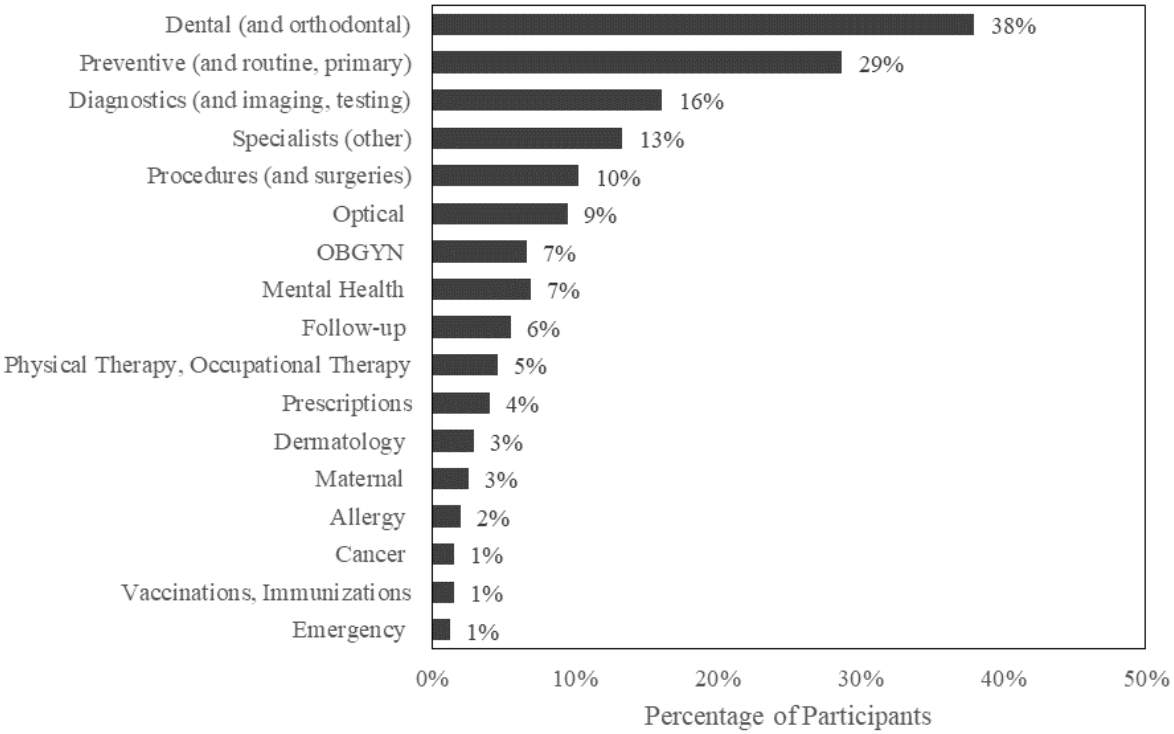
Table 1. Sample characteristics (N=2570).

Variable	Value
Numeric variables, mean (SD)	
Age (years)	37.31 (12.63)
Distance to hospital (miles)	2.30 (0.95)
Health-related worry (scale of 0-100)	52.27 (26.71)
COVID-19–related worry (scale of 0-100)	74.93 (20.54)
Categorical variables, n (%)	
Gender	
Cisgender women	2456 (95.6)
Cisgender men	324 (12.6)
Transgender or nonbinary individuals	118 (5.6)
Sexual identity	
Heterosexual	1680 (65.4)
Sexual minority	1202 (46.8)
Race and ethnicity	
White	2464 (95.9)
Black	48 (1.9)
Latina, Latino, Latinx, or Hispanic	174 (6.8)
American Indian or Alaska Native	0 (0)
Asian or Pacific Islander	199 (7.7)
Biracial	7 (<1)
Education	
High school or less	57 (2.2)
Less than a Bachelor's degree	419 (16.3)
Bachelor's degree	783 (30.5)
Enrolled in graduate school	287 (11.2)
Graduate or professional school	894 (34.8)
Doctorate degree (eg, MD, PhD, DrPH)	449 (17.5)

Almost half (1227/2570, 47.8%) of all participants reported experiencing delays or interruptions in their health care. Of these, 8.3% (102/1227) reported switching their care to telemedicine, and 12.6% (154/1227) reported experiencing health care access issues (eg, barriers associated with insurance coverage). Most participants (921/1227, 75.1%) who experienced delays specified the types of affected care. Some individuals solely responded with phrases such as “doctor’s appointment,” which resulted in no associated code due to the lack of the specificity of the response. Among those that specified the type of care, the three most commonly reported types of care affected included dental (351/921, 38.1%), preventive (269/921, 29.2%), and diagnostics (151/921, 16.4%) care (Figure 1). Emergency care was the least commonly reported type of care affected (12/921, 1.3%) among this sample, followed by vaccinations (14/921, 1.5%) and cancer care (14/921, 1.5%). Further, 10.2% (94/921) of the participants who responded to this item reported disruptions to the health care of others (eg, child, spouse, and parent; not included in Figure 1).

XSL•FO
RenderX

Figure 1. Types of health care impacted by the COVID-19 pandemic (n=921). Proportions do not sum to 100% because participants may have listed >1 type of care. OGBYN: obstetrics and gynecology.



In Table 2, we present the reported barriers to receiving health care services, which were based on follow-up questions that were sent to individuals who indicated an interest in participating in follow-up questionnaires. The top 3 barriers to care—each reported by about one-third of participants—included a fear of SARS-CoV-2 infection (126/374, 33.7%), provider discouragement (122/374, 32.6%), and the feeling that their health care concerns were not as important as others' concerns (118/374, 31.6%).

Table 2. Barriers to receiving health care during the COVID-19 pandemic (n=1025). Responses to the following question are presented: "Since the start of the coronavirus pandemic, have you needed to get healthcare but haven't gotten it? Why?"

Responses	Value, n (%)[a]
No	651 (63.5)
Yes	374 (36.5)
I was afraid of getting infected	126 (33.7)
My health care provider discouraged me from coming in	122 (32.6)
I felt like my concern/need wasn't as important as other people's	118 (31.6)
My health care provider is unavailable	108 (28.9)
I felt like my symptoms weren't severe enough	89 (23.8)
I have no health insurance	15 (4)
I cannot afford my copay or deductible	14 (3.7)
Other	60 (16)

[a]Proportions do not sum to 100% because participants may have listed >1 reason.

Determinants of Patient-Reported Health Care Delays

Table 3 shows the associations between the dependent variable "any health care delays" and independent variables as a series of separate bivariate tests. Several demographic factors were associated with reporting any health care delays resulting from the pandemic. On average, older age, cisgender women, White individuals, higher levels of education, higher levels of self-reported worries about general health, and higher levels of self-reported worries about COVID-19 were all associated with experiencing health care delays. We then entered all independent variables, regardless of their significance at the bivariate level, into a multiple logistic regression model. The results from the logistic regression models indicated that age (P<.001), gender

XSL•FO
RenderX

identity (*P*<.001), race (*P*=.005), education (*P*=.007), and the degree of worry about COVID-19 (*P*<.001) and general health (*P*<.001) were all significantly associated with experiencing health care delays due to the pandemic. The largest effect was observed for gender; the odds of patient-reported health care delays for cisgender men were 0.60 (95% CI 0.46-0.78) times that of cisgender women.

Table 3. Bivariate analyses[a] of the reasons for experiencing health care delays resulting from the COVID-19 pandemic (N=2570).

Independent variables	Yes	No	*P* value
Age (years), mean (SD)	38.9 (13.1)	36.0 (12.1)	<.001
Gender identity, %[b]			
Cisgender women	87.2	83.8	<.001
Cisgender men	8.3	13.4	—[c]
Transgender or nonbinary individuals	4.6	3.8	—
Race, %[b]			
White	87.8	83.9	.005
People of color	12.2	16.1	—
Sexual identity, %[b]			
Heterosexual	60	57.6	.22
Sexual minority	40	42.4	—
Education, %[b]			
Less than a Bachelor's degree	17	14.6	.007
Bachelor's degree or some graduate school	38.9	35.3	—
Graduate or doctorate degree	44.1	50.2	—
Distance to nearest hospital (miles), mean (SD)	2.3 (1.0)	2.3 (0.9)	.15
Worry about general health (scale of 0-100), mean (SD)	54.5 (27.0)	50.1 (26.2)	<.001
Worry about COVID-19 (scale of 0-100), mean (SD)	76.3 (20.4)	73.6 (20.6)	<.001

[a]Bivariate tests are used to determine whether each independent variable is related to each dependent variable. Continuous independent variables are used in an independent samples *t* test to identify a significant association, while categorical independent variables are used in a chi-square test of association.

[b]Percentages are based on column totals.

[c]Not available.

We also examined delays in specific types of health care—dental, preventive, and diagnostics care. Both the bivariate tests (Table 4) and the multiple logistic regression model (Table 5) showed that the only significant independent variable that was associated with dental delays was self-reported worries about general health. Increased levels of worry were associated with fewer delays in dental care (*P*=.002). Our bivariate analyses identified two factors that were associated with experiencing delays in diagnostics (Table 4)—age and self-reported worry about general health. In the multiple logistic regression model, both factors remained significant at the .05 level of confidence. Age had a larger effect on experiencing delays in diagnostics. Every 10-year increase in age increased the odds of experiencing delays in diagnostics by a multiple of 1.36. The bivariate tests of association for delays in preventive care showed that age (*P*=.03) and race and ethnicity (*P*=.02) were significant. Neither of these factors remained significant in the multiple logistic regression model, and although there was trend between race and delays (people of color experienced fewer delays than White participants), it was not significant (*P*=.06).

XSL•FO
RenderX

Table 4. Bivariate analyses of the reasons for experiencing health care delays resulting from the COVID-19 pandemic among those who reported experiencing any health care delays (N=921).

Independent variables	Dependent variables								
	Delays in dental care			Delays in diagnostics			Delays in preventive care		
	Yes	No	*P* value	Yes	No	*P* value	Yes	No	*P* value
Age (years), mean (SD)	39.3 (13.3)	39.3 (13.3)	.50	43.6 (13.3)	38.8 (13.1)	<.001	41.1 (13.3)	38.9 (13.0)	.03
Gender, %			.94			.36			.24
Cisgender women	86.6	87		88.1	86.6		89.2	85.9	
Cisgender men	8.8	8.3		6	9		7.8	8.7	
Transgender or nonbinary individuals	4.6	4.7		6	4.4		3	5.4	
Race, %			.99			.38			.02
White	88.9	89.1		86.7	89.5		92.9	87.4	
People of color	11.1	10.9		13.3	10.5		7.1	12.6	
Sexual identity, %			.22			.99			.07
Heterosexual	61.1	56.9		58.3	58.5		63.2	56.5	
Sexual minority	38.9	43.1		41.7	41.5		36.8	43.5	
Education, %			.89			.60			.09
Less than a Bachelor's degree	13.7	14.7		14.6	14.3		11.9	15.3	
Bachelor's degree or some graduate school	35.9	34.9		31.8	36		32.3	36.5	
Graduate or doctorate degree	50.4	50.4		53.6	49.7		55.8	48.2	
Distance miles to nearest hospital (miles), mean (SD)	2.3 (0.9)	2.4 (1.0)	.25	2.3 (1.0)	2.3 (1.0)	.87	2.3 (0.9)	2.3 (1.0)	.73
Worry about general health (scale of 0-100), mean (SD)	50.5 (27.2)	56.1 (26.4)	.002	60.2 (26.0)	52.8 (26.8)	.001	54.3 (25.2)	53.8 (27.5)	.81
Worry about COVID (scale of 0-100), mean (SD)	75.8 (20.6)	77.1 (20.3)	.38	79.0 (19.3)	76.1 (20.6)	.11	77.2 (19.2)	76.3 (20.9)	.54

XSL•FO
RenderX

Table 5. Multiple logistic regression of reasons for experiencing health care delays resulting from the COVID-19 pandemic.

Independent variables[a]	Dependent variables[b]							
	Any delays		Delays in dental care		Delays in diagnostics		Delays in preventive care	
	P value	95% CI	*P* value	95% CI	*P* value	95% CI	*P* value	95% CI
Age (years; 10-year increments)	<.001	1.12-1.28	.25	0.95-1.19	<.001	1.15-1.53	.18	0.96-1.22
Gender								
Cisgender men	<.001	0.46-0.78	.93	0.60-1.59	.17	0.27-1.20	.60	0.50-1.46
Transgender or nonbinary individuals	.36	0.80-1.81	.75	0.56-2.16	.48	0.55-3.03	.39	0.29-1.52
Race								
People of color	.07	0.64-1.02	.70	0.70-1.68	.10	0.89-2.69	.05	0.34-0.99
Sexual identity								
Sexual minority	.01	1.04-1.46	.28	0.63-1.14	.31	0.82-1.81	.26	0.61-1.14
Education								
Bachelor's degree or some graduate school	.43	0.87-1.40	.63	0.72-1.71	.87	0.60-1.90	.32	0.79-2.09
Graduate or doctorate degree	.02	1.03-1.64	.85	0.69-1.57	.58	0.69-2.03	.07	0.96-2.39
Distance (miles) to nearest hospital	.34	0.96-1.13	.31	0.81-1.07	.68	0.80-1.16	.70	0.89-1.19
Worry about general health (10-point increments)	.01	1.01-1.08	.002	0.86-0.97	.02	1.01-1.19	.95	0.94-1.06
Worry about COVID (10-point inccrements)	.88	0.96-1.05	.35	0.96-1.12	.62	0.88-1.09	.72	0.93-1.10
Intercept	<.001	0.17-0.44	.41	0.31-1.62	<.001	0.01-0.12	<.001	0.09-0.53

[a]The reference group for all categorical variables are as follows: cisgender woman (gender), White (race), heterosexual (sexual identity), and less than Bachelor's degree (education).

[b]There were 2570 observations for any delays, 907 observations for delays in dental care, 907 observations for delays in diagnostics, and 907 observations for delays in preventive care.

Discussion

Principal Findings

In June 2020, Dr Lasic, an interventional cardiologist at Jamaica Hospital Medical Center and Lenox Hill Hospital in New York, made the following prediction: "I think the toll on non-COVID patients will be much greater than COVID deaths" [18]. To our knowledge, ours is the only study that aimed at understanding the toll on patients without COVID-19 by describing rates of multiple types of health care delays as well as characterizing people who were the most affected by these delays during the beginning of the pandemic. In June 2020, under dynamic circumstances, researchers found that 32% and 12% of adults deliberately delayed or avoided routine care and emergency care, respectively [19]. Our findings suggest that a not insignificant proportion of the US populace may be experiencing delayed or interrupted care as a result of the pandemic, which may lead to decrements in health over the long term. Across all types of delays, older people, people with higher levels of education, and people who were more worried about their health were more likely to report delays or interruptions in their care. Men and people of color were less likely to report delays or interruptions.

Our findings highlight that dental care appears to be the most impacted (reported by over one-third of participants; 351/921, 38.1%). This is unsurprising, given that many dental offices closed at the outset of the pandemic due to concerns of patient and provider safety. However, studies have shown a relationship between dental health and heart disease, between dental health and diabetes, and between dental health and prenatal outcomes [20-22]. This is suggestive of the potential, downstream, deleterious health implications of delayed or interrupted dental care. Dentists are often also on the frontlines of identifying child abuse [23] and intimate partner violence [24], of which both have reportedly increased in incidence during the pandemic [25-27]. Further, a recent study conducted in Qatar has suggested that routine oral care is associated with a reduction in the risk of COVID-19 complications (eg, hospitalization and ventilation) [28]. Several months into the pandemic (July 2020), the American Dental Association released a policy declaring that dentistry was essential care (ie, care that is integral to systemic health) [29]. Of great concern are our findings associated with the frequency of reported delays in preventive and diagnostic care, particularly delays in such care for older adults. The implications of this finding are described in the *Introduction* section.

The lessons learned from the initial surge of COVID-19 cases can help mitigate potential future disruptions. Mitigation strategies for future disruptions include effective patient prioritization and triage [30-34]. For instance, Medically Necessary Time-Sensitive Prioritization is a scoring system for surgical triage that integrates multiple factors associated with

XSL•FO
RenderX

the environment, patient risk, and ethics that are intended to be generalizable across hospital settings [35]. Additionally, the Centers for Medicare & Medicaid Services have released guidelines for the provision of non–COVID-19 care [36-38], However, considerations should also include patient life factors (eg, family support and transportation) as well as systems factors at the hospital (eg, visitation policy), local and state (eg, infection control mandates), or national levels (eg, policies). Using systems thinking perspectives is necessary for addressing complex problems [39]. Cancer care providers have also recommended balancing delays against COVID-19 risk, practicing effective social distancing, and managing the appropriate allocation of resources [32]. Key to reducing delays among patients is the relaying of accessible, clear, and actionable messaging by health care systems about how and when to safely access health care.

More research on health care delays and interruptions during the pandemic is needed. Further investigation is needed into the differences between the care of chronic illness and emergent conditions. For instance, the first pandemic surge resulted in a decline in hospitalization rates and lengths of hospital stay associated with acute cardiovascular conditions as well as other common conditions, such as acute appendicitis, bone fractures, cancer, and live births. Given the evidence on differences based on sex [40], race and ethnicity [41], sexual and gender identity [42,43], and other key factors, we need to investigate sociodemographic differences in delayed or interrupted care. Further, the significant economic effects of the COVID-19 pandemic as well as the high rates of job loss may result in the loss of insurance. Thus, more work is needed to understand how these factors affect health care. The potential impacts on patient-level outcomes such as mental health, quality of life, and pain must also be considered.

Study Limitations

There are some key limitations that need to be considered. First, this study relied on a convenience sample that is not representative in terms of race, ethnicity, education, age, and gender. Convenience samples are limited in terms of the generalizability of prevalence rates. However, they have the advantage of efficiency, which is critical in a dynamic situation, and can be useful for understanding patterns and relationships associated with a phenomenon of interest. Recent research has shown that internet-based convenience samples correlate with truly random probability samples to a surprising extent. For example, Mullinix et al [44] found a correlation of 0.75 when they conducted 20 experiments that replicated results based on national probability samples on Amazon Mechanical Turk (an internet-based survey platform). More recently, Coppock [45] conducted 15 replication experiments and similarly found that convenience and national probability samples provided similar estimates of treatment and moderator effects. However, despite the limitation of using a convenience sample, our study is unique and critical to documenting experiences at the outset of the pandemic, given that the conditions under which the data were collected are not replicable. In addition, a study conducted over the internet does not reach individuals who do not have access to the internet. However, this is perhaps the only mechanism for capturing data from a fleeting moment in time that is characterized by dynamic and uncertain circumstances.

Conclusions

The pandemic has amplified people's attention to health disparities and inequities in the United States and has created a need for inclusive and equitable research. The limitations of this study highlight pervasive gaps and challenges associated with conducting such research (and doing so efficiently, given the dynamic nature of the COVID-19 pandemic). There is an urgent need for sharing lessons learned, disseminating effective strategies for reaching more diverse populations (eg, engaging leaders of marginalized communities, understanding and addressing research hesitancy, etc), and encouraging the research community to use and improve upon these strategies in future research. Although the conditions for this study cannot be replicated, the methodological lessons learned can serve as a sort of pilot study for future crises, thereby creating more diverse and inclusive bodies of research that drive health equity forward. Without explicit discussions of research limitations, the research community cannot make progress in collecting data to inform the design of effective programs for addressing health inequities that have existed long before the COVID-19 health crisis. The impact of health care disruptions resulting from the COVID-19 pandemic may be difficult to measure in the short term. However, characterizing these disruptions and improving research methods for such characterizations is critical to informing systemic and equitable planning, mitigation, and recovery strategies for the long term.

Acknowledgments

The open-access publication fee for this article was supported by the Research Open Access Publishing Fund of the University of Illinois at Chicago.

Conflicts of Interest

None declared.

References

1. Flume P, Gray S, Bowman CM, Kerrigan C, Lester M, Virella-Lowell I. Emergency preparedness for the chronically ill. Am J Nurs 2005 Mar;105(3):68-72. [doi: 10.1097/00000446-200503000-00028] [Medline: 15729049]
2. Ford ES, Mokdad AH, Link MW, Garvin WS, McGuire LC, Jiles RB, et al. Chronic disease in health emergencies: in the eye of the hurricane. Prev Chronic Dis 2006 Apr;3(2):A46 [FREE Full text] [Medline: 16539787]

XSL•FO
RenderX

3. Ryan BJ, Franklin RC, Burkle FM, Aitken P, Smith E, Watt K, et al. Reducing disaster exacerbated non-communicable diseases through public health infrastructure resilience: Perspectives of Australian disaster service providers. PLoS Curr 2016 Dec 21;8:1-26 [FREE Full text] [doi: 10.1371/currents.dis.d142f36b6f5eeca806d95266b20fed1f] [Medline: 28239511]
4. Slama S, Kim HJ, Roglic G, Boulle P, Hering H, Varghese C, et al. Care of non-communicable diseases in emergencies. Lancet 2017 Jan 21;389(10066):326-330. [doi: 10.1016/S0140-6736(16)31404-0] [Medline: 27637675]
5. Papautsky EL, Hamlish T. Patient-reported treatment delays in breast cancer care during the COVID-19 pandemic. Breast Cancer Res Treat 2020 Nov;184(1):249-254 [FREE Full text] [doi: 10.1007/s10549-020-05828-7] [Medline: 32772225]
6. Farley TA, Dalal MA, Mostashari F, Frieden TR. Deaths preventable in the U.S. by improvements in use of clinical preventive services. Am J Prev Med 2010 Jun;38(6):600-609. [doi: 10.1016/j.amepre.2010.02.016] [Medline: 20494236]
7. Maciosek MV, Coffield AB, Flottemesch TJ, Edwards NM, Solberg LI. Greater use of preventive services in U.S. health care could save lives at little or no cost. Health Aff (Millwood) 2010 Sep;29(9):1656-1660. [doi: 10.1377/hlthaff.2008.0701] [Medline: 20820022]
8. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003 Jun 26;348(26):2635-2645. [doi: 10.1056/NEJMsa022615] [Medline: 12826639]
9. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol 2012 Sep 11;22(17):R741-R752 [FREE Full text] [doi: 10.1016/j.cub.2012.07.024] [Medline: 22975005]
10. Werner MT, Carey RM, Albergotti WG, Lukens JN, Brody RM. Impact of the COVID-19 Pandemic on the Management of Head and Neck Malignancies. Otolaryngol Head Neck Surg 2020 Jun;162(6):816-817. [doi: 10.1177/0194599820921413] [Medline: 32312163]
11. Stensland KD, Morgan TM, Moinzadeh A, Lee CT, Briganti A, Catto JWF, et al. Considerations in the triage of urologic surgeries during the COVID-19 pandemic. Eur Urol 2020 Jun;77(6):663-666 [FREE Full text] [doi: 10.1016/j.eururo.2020.03.027] [Medline: 32279903]
12. Reza N, DeFilippis EM, Jessup M. Secondary impact of the COVID-19 pandemic on patients with heart failure. Circ Heart Fail 2020 May;13(5):e007219. [doi: 10.1161/CIRCHEARTFAILURE.120.007219] [Medline: 32352841]
13. Ding YY, Ramakrishna S, Long AH, Phillips CA, Montiel-Esparza R, Diorio CJ, et al. Delayed cancer diagnoses and high mortality in children during the COVID-19 pandemic. Pediatr Blood Cancer 2020 Sep;67(9):e28427 [FREE Full text] [doi: 10.1002/pbc.28427] [Medline: 32588960]
14. Patt D, Gordan L, Diaz M, Okon T, Grady L, Harmison M, et al. Impact of COVID-19 on cancer care: How the pandemic Is delaying cancer diagnosis and treatment for American seniors. JCO Clin Cancer Inform 2020 Nov;4:1059-1071 [FREE Full text] [doi: 10.1200/CCI.20.00134] [Medline: 33253013]
15. Braun V, Clark V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology: Vol. 2. Research Designs: Quantitative, qualitative, neuropsychological, and biological. Washington, DC, US: American Psychological Association; 2012:57-71.
16. Ryan GW, Bernard HR. Data management and analysis methods. In: Denzin NK, Lincoln YS, editors. Handbook of Qualitative Research, 2nd Edition. Thousand Oaks: Sage Publications; 2000:769-802.
17. The R Project for Statistical Computing. The R Foundation. URL: https://www.r-project.org/ [accessed 2021-05-05]
18. Rosenbaum L. The untold toll - The pandemic's effects on patients without Covid-19. N Engl J Med 2020 Jun 11;382(24):2368-2371. [doi: 10.1056/NEJMms2009984] [Medline: 32302076]
19. Czeisler ME, Marynak K, Clarke KEN, 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]
20. Khader YS, Ta'ani Q. Periodontal diseases and the risk of preterm birth and low birth weight: a meta-analysis. J Periodontol 2005 Feb;76(2):161-165. [doi: 10.1902/jop.2005.76.2.161] [Medline: 15974837]
21. Sanz M, Ceriello A, Buysschaert M, Chapple I, Demmer RT, Graziani F, et al. Scientific evidence on the links between periodontal diseases and diabetes: Consensus report and guidelines of the joint workshop on periodontal diseases and diabetes by the International diabetes Federation and the European Federation of Periodontology. Diabetes Res Clin Pract 2018 Mar;137:231-241. [doi: 10.1016/j.diabres.2017.12.001] [Medline: 29208508]
22. Yu YH, Chasman DI, Buring JE, Rose L, Ridker PM. Cardiovascular risks associated with incident and prevalent periodontal disease. J Clin Periodontol 2015 Jan;42(1):21-28 [FREE Full text] [doi: 10.1111/jcpe.12335] [Medline: 25385537]
23. Smitt HS, de Leeuw J, de Vries T. Association between severe dental caries and child abuse and neglect. J Oral Maxillofac Surg 2017 Nov;75(11):2304-2306. [doi: 10.1016/j.joms.2017.05.004] [Medline: 28586637]
24. Lea SJ, Quinn B, Reynolds PA. The role and education of dental care professionals in identifying domestic violence: Report of an audience participation exercise and round table discussion. Technology, Knowledge and Learning 2016 Dec 2;22:219-226 [FREE Full text] [doi: 10.1007/s10758-016-9293-9]
25. Bettinger-Lopez C, Bro A. A double pandemic: Domestic violence in the age of COVID-19. Council on Foreign Regulations. 2020 May 13. URL: https://www.cfr.org/in-brief/double-pandemic-domestic-violence-age-covid-19 [accessed 2021-05-06]
26. Brown SM, Doom JR, Lechuga-Peña S, Watamura SE, Koppels T. Stress and parenting during the global COVID-19 pandemic. Child Abuse Negl 2020 Dec;110(Pt 2):104699 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104699] [Medline: 32859394]

XSL•FO
RenderX

27. Velonis A, McGown M. When home does not offer shelter: partner violence in the time of pandemics. University of Illinois Chicago. 2020 Apr 04. URL: https://publichealth.uic.edu/news-stories/when-home-does-not-offer-shelter-partner-violence-in-the-time-of-pandemics/ [accessed 2021-05-06]
28. Marouf N, Cai W, Said KN, Daas H, Diab H, Chinta V, et al. Association between periodontitis and severity of COVID-19 infection: A case-control study. J Clin Periodontol 2021 Apr;48(4):483-491 [FREE Full text] [doi: 10.1111/jcpe.13435] [Medline: 33527378]
29. Current policies. American Dental Association. URL: https://www.ada.org/en/advocacy/current-policies#publichealthemergencies [accessed 2020-09-16]
30. Brody RM, Albergotti WG, Shimunov D, Nicolli E, Patel UA, Harris BN, et al. Changes in head and neck oncologic practice during the COVID-19 pandemic. Head Neck 2020 Jul;42(7):1448-1453 [FREE Full text] [doi: 10.1002/hed.26233] [Medline: 32357380]
31. Hanna TP, Evans GA, Booth CM. Cancer, COVID-19 and the precautionary principle: prioritizing treatment during a global pandemic. Nat Rev Clin Oncol 2020 May;17(5):268-270 [FREE Full text] [doi: 10.1038/s41571-020-0362-6] [Medline: 32242095]
32. Kutikov A, Weinberg DS, Edelman MJ, Horwitz EM, Uzzo RG, Fisher RI. A war on two fronts: Cancer care in the time of COVID-19. Ann Intern Med 2020 Jun 02;172(11):756-758 [FREE Full text] [doi: 10.7326/M20-1133] [Medline: 32219410]
33. Nagar H, Formenti SC. Cancer and COVID-19 - potentially deleterious effects of delaying radiotherapy. Nat Rev Clin Oncol 2020 Jun;17(6):332-334 [FREE Full text] [doi: 10.1038/s41571-020-0375-1] [Medline: 32341524]
34. O'Leary MP, Choong KC, Thornblade LW, Fakih MG, Fong Y, Kaiser AM. Management considerations for the surgical treatment of colorectal cancer during the global Covid-19 pandemic. Ann Surg 2020 Aug;272(2):e98-e105 [FREE Full text] [doi: 10.1097/SLA.0000000000004029] [Medline: 32675510]
35. Prachand VN, Milner R, Angelos P, Posner MC, Fung JJ, Agrawal N, et al. Medically necessary, time-sensitive procedures: Scoring system to ethically and efficiently manage resource scarcity and provider risk during the COVID-19 pandemic. J Am Coll Surg 2020 Aug;231(2):281-288 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2020.04.011] [Medline: 32278725]
36. Helping private practices navigate non-essential care during COVID-19. American Medical Association. URL: https://www.ama-assn.org/delivering-care/public-health/helping-private-practices-navigate-non-essential-care-during-covid-19 [accessed 2020-09-18]
37. Healthcare facilities: Managing operations during the COVID-19 pandemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html [accessed 2020-09-18]
38. Coronavirus (COVID-19) partner resources. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/outreach-education/partner-resources/coronavirus-covid-19-partner-toolkit [accessed 2020-09-18]
39. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
40. Veldhuis CB, Maki P, Molina K. Psychological and neighborhood factors associated with urban women's preventive care use. J Behav Med 2020 Jun;43(3):346-364 [FREE Full text] [doi: 10.1007/s10865-019-00122-4] [Medline: 31865485]
41. Phillips KA, Mayer ML, Aday LA. Barriers to care among racial/ethnic groups under managed care. Health Aff (Millwood) 2000;19(4):65-75. [doi: 10.1377/hlthaff.19.4.65] [Medline: 10916961]
42. Hsieh N, Ruther M. Despite increased insurance coverage, nonwhite sexual minorities still experience disparities in access to care. Health Aff (Millwood) 2017 Oct 01;36(10):1786-1794. [doi: 10.1377/hlthaff.2017.0455] [Medline: 28971924]
43. Progovac AM, Lê Cook B, Mullin BO, McDowell A, Sanchez R MJ, Wang Y, et al. Identifying gender minority patients' health and health care needs in administrative claims data. Health Aff (Millwood) 2018 Mar;37(3):413-420 [FREE Full text] [doi: 10.1377/hlthaff.2017.1295] [Medline: 29505378]
44. Mullinix KJ, Leeper TJ, Druckman JN, Freese J. The generalizability of survey experiments. Journal of Experimental Political Science 2016 Jan 12;2(2):109-138. [doi: 10.1017/xps.2015.19]
45. Coppock A. Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. Political Science and Research Methods 2018 Mar 27;7(3):613-628 [FREE Full text] [doi: 10.1017/psrm.2018.10]

XSL•FO
RenderX

Edited by R Kukafka, C Basch; submitted 02.11.20; peer-reviewed by M Koo, J Ziegenfuss, M Callegaro; comments to author 14.01.21; revised version received 16.03.21; accepted 20.04.21; published 19.05.21.

Please cite as:
Papautsky EL, Rice DR, Ghoneima H, McKowen ALW, Anderson N, Wootton AR, Veldhuis C
Characterizing Health Care Delays and Interruptions in the United States During the COVID-19 Pandemic: Internet-Based, Cross-sectional Survey Study
J Med Internet Res 2021;23(5):e25446
URL: https://www.jmir.org/2021/5/e25446
doi:10.2196/25446
PMID:33886489

©Elizabeth Lerner Papautsky, Dylan R Rice, Hana Ghoneima, Anna Laura W McKowen, Nicholas Anderson, Angie R Wootton, Cindy Veldhuis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2021. 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

Short Paper

Telehealth Use by Age and Race at a Single Academic Medical Center During the COVID-19 Pandemic: Retrospective Cohort Study

Jennifer P Stevens[1], MD, MS; Oren Mechanic[2], MD; Lawrence Markson[3], MD; Ashley O'Donoghue[1], PhD; Alexa B Kimball[2], MD

[1]Center for Healthcare Delivery Science, Beth Israel Deaconess Medical Center, Boston, MA, United States

[2]Harvard Medical Faculty Physicians, Beth Israel Deaconess Medical Center, Boston, MA, United States

[3]Information Systems, Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:
Jennifer P Stevens, MD, MS
Center for Healthcare Delivery Science
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Boston, MA, 02215
United States
Phone: 1 6176677165
Email: jpsteven@Bidmc.harvard.edu

Abstract

Background: During the COVID-19 pandemic, many ambulatory clinics transitioned to telehealth, but it remains unknown how this may have exacerbated inequitable access to care.

Objective: Given the potential barriers faced by different populations, we investigated whether telehealth use is consistent and equitable across age, race, and gender.

Methods: Our retrospective cohort study of outpatient visits was conducted between March 2 and June 10, 2020, compared with the same time period in 2019, at a single academic health center in Boston, Massachusetts. Visits were divided into in-person visits and telehealth visits and then compared by racial designation, gender, and age.

Results: At our academic medical center, using a retrospective cohort analysis of ambulatory care delivered between March 2 and June 10, 2020, we found that over half (57.6%) of all visits were telehealth visits, and both Black and White patients accessed telehealth more than Asian patients.

Conclusions: Our findings indicate that the rapid implementation of telehealth does not follow prior patterns of health care disparities.

(J Med Internet Res 2021;23(5):e23905) doi:10.2196/23905

KEYWORDS

access; barrier; cohort; COVID-19; demographic; equity of care; equity; outpatient; telehealth

Introduction

During the COVID-19 pandemic, many hospitals reduced in-person ambulatory care. Through payment parity and reductions in administrative barriers, including the Health Insurance Portability and Accountability Act of 1996 waivers by Health and Human Services, physicians rapidly began providing care through telemedicine [1,2].

Our large academic physicians' group in Boston, Massachusetts, similarly expanded telehealth in mid-March 2020. This sudden move, however, risked exacerbating care inequity across racial and socioeconomic groups [3,4]. Given potential barriers faced by different populations [5], we investigated whether telehealth use was consistent and equitable across age, race, and gender.

XSL•FO
RenderX

Methods

Study Population and Data Source

This study was deemed exempt by the institutional review board at the Beth Israel Deaconess Medical Center. In-person and telehealth visits to outpatient clinics staffed by our practicing faculty, between March 2 and June 10, 2020, for patients aged ≥18 years were identified and compared to visits to identical clinics during the same period in 2019.

Study Variables

Patient age (in deciles), self-identified race, and gender were extracted from the electronic health records.

Statistical Analysis

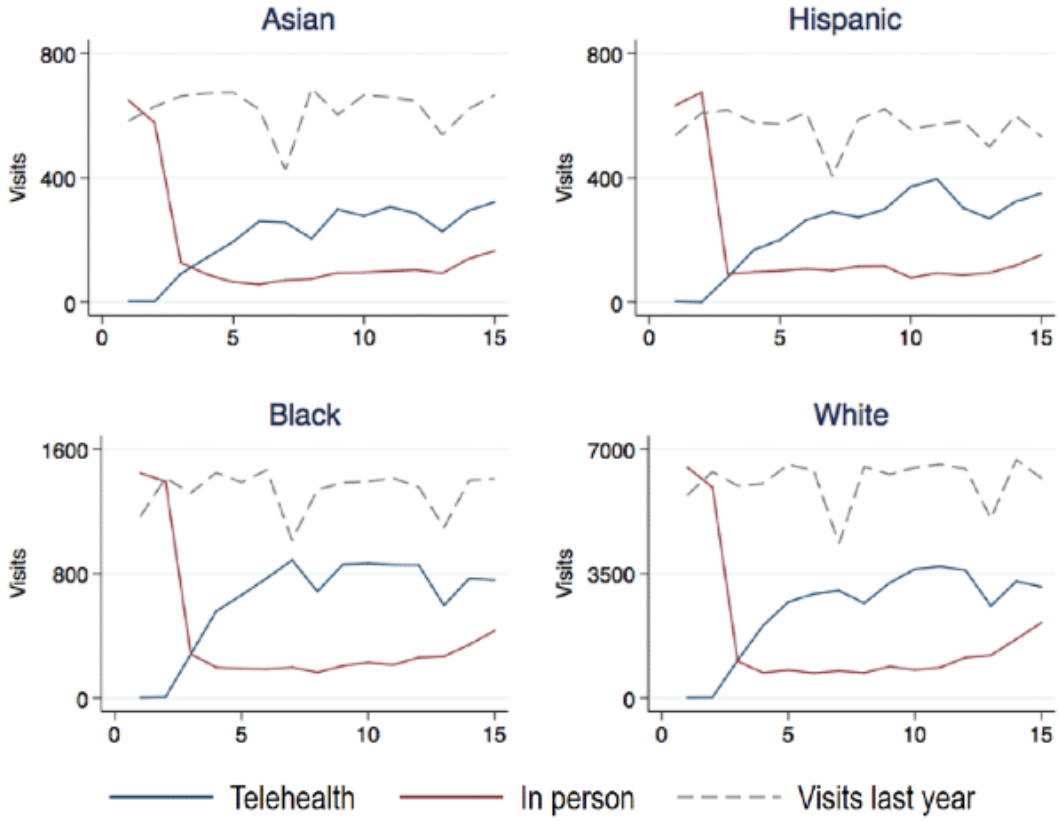
Statistical tests were performed using Stata SE (version 14.2, StataCorp). In-person and telehealth use were stratified by week and race and compared to the same period in 2019. We conducted similar comparisons by gender and age deciles and explored interaction terms to determine whether different groups of age deciles and race used telehealth more or less than other groups. We further conducted a subgroup analysis of care provided after April 27; after this date, we had more complete documentation of whether the telehealth visit was conducted using video technology or over the telephone. In this subgroup, we explored whether older adults (≥65 years old) made use of video technology at similar rates as younger patients. We similarly conducted a subgroup analysis during this period to investigate whether non-White patients used video technology at similar rates as White patients.

Results

Between March 2 and June 10, 2020, a total of 129,844 ambulatory visits were conducted, compared to 180,831 visits during the same period in 2019, which indicates a reduction of 28%. Compared to 2019, visits in 2020 decreased for all racial groups (White patients: 31%, Black patients: 23%, Hispanic patients: 26%, and Asian patients: 39%; *P*<.001).

Among visits in 2020, a total of 74,846 (57.6%) visits were conducted through telehealth. Overall, Black (n=9414 of 15,423, 61%) and White (n=37,620 of 63,397, 59.3%) patients used telehealth rather than in-person visits at higher rates; Asian patients used telehealth the least (n=3162 of 5661, 55.9%) (Figure 1). Patients with "unknown" racial designations displayed the lowest rates of telehealth use (*P*<.001).

Figure 1. Comparison of outpatient visits in 2020 with those conducted during the same time period in 2019 (dotted line), by race at a single academic health system in Boston, Massachusetts. While all racial groups experienced decreased access to outpatient care owing to the surge in COVID-19 cases from March to June 2020, Asian patients had reduced access to any care (in-person and telehealth) and a decreased rate of access to telehealth services than those of other racial groups. The x-axis represents weeks since March 2020.



Older patients accessed more outpatient care than younger patients (*P*<.001) (Table 1). Differences were observed by age and race; differences were most prominent among Asian patients (*P*<.001). Asian and Hispanic patients aged under 65 years used telehealth less than age-matched White patients (*P*<.001), but there was no difference in telehealth usage between Black and White patients under 65 years of age (*P*=.07).

XSL•FO
RenderX

In a subgroup analysis conducted among patients who sought care after April 27, we found that older adults were less likely to use video technology (19% of telehealth visits) rather than the telephone compared to their younger counterparts (26% of telehealth visits; $P<.001$). White patients used video technology significantly more (25% of telehealth visits) than those of all other races (Black patients: 16%, Hispanic patients: 23%, and Asian patients: 17%; $P<.001$).

Table 1. Overall usage of in-person and telehealth services from March 2 to June 10, 2020, at a single academic medical center in Boston, Massachusetts, among patients stratified by age and race, recorded at the time of registration (N=129,844).

Parameters	Total	In-person visits	Telehealth visits	*P* value
Age (years), n (%)				<.001
18-29	11,121 (100)	5179 (46.6)	5942 (53.4)	
30-39	19,675 (100)	10,234 (52.0)	9441 (48.0)	
40-49	15,831 (100)	6623 (41.8)	9208 (58.2)	
50-59	23,238 (100)	9437 (40.6)	13,801 (59.4)	
60-69	28,120 (100)	11,535 (41.0)	16,585 (59.0)	
70-79	20,972 (100)	7914 (37.7)	13,058 (62.3)	
≥80	10,887 (100)	4076 (37.4)	6811 (62.6)	
Race, n (%)				<.001
Asian	5661 (100)	2499 (44.1)	3162 (55.9)	
Black	15,423 (100)	6009 (39.0)	9414 (61.0)	
Hispanic	6243 (100)	2656 (42.5)	3587 (57.5)	
Other	8499 (100)	3817 (44.9)	4692 (55.1)	
Unknown	30,621 (100)	14,240 (46.5)	16,381 (53.5)	
White	63,397 (100)	25,777 (40.7)	37,620 (59.3)	
Total, n (%)	129, 844 (100)	54,998 (42.4)	74,846 (57.6)	

[a]N/A: not applicable.

Discussion

Principal Findings

Our results indicate that implementation of telehealth does not necessarily exacerbate the inequity in health care access, but it should be monitored carefully. Patients had fewer ambulatory visits of any kind during the COVID-19 pandemic; however, Asian and White patients accessed care less than Black and Hispanic patients compared to baseline. Black and White patients accessed telehealth care more than Hispanic and Asian patients. Patients with an "unknown" racial or ethnic designation upon registration had the least access to care.

Contrary to concerns that older patients might have difficulty navigating technology [6], our older patients used telehealth more often than younger patients. The potential reasons for this could be that younger patients perceived their clinical needs to be less urgent, had variable awareness of telehealth services, or experienced additional barriers [7]. However, we found that among telehealth users, patients over 65 years of age were less likely to use video technology, which may reflect concerns with technology; these concerns have been explored by other investigators [6,8]. Similarly, Black, Asian, and Hispanic patients were less likely to use video technology than White patients.

Limitations

The limitations of this study include the use of administrative data and self-reported racial designations; approximately 22% of patients in our study belonged to an "unknown" racial category, which may limit our ability to draw inferences [9,10]. Further, our documentation of telehealth early in the pandemic was unable to distinguish telephonic and video visits, which may vary across racial and age groups [6].

Conclusions

In conclusion, after a rapid increase in telehealth use at a single academic medical center in Boston, Massachusetts, we observed variable engagement of our patient population by both race and age in telehealth, but the trends did not mirror previously described patterns of health access disparity.

Acknowledgments

JPS is supported by award# K08HS024288 from the Agency for Healthcare Research and Quality. The content of this paper is solely the responsibility of the authors and does not represent the official views of the Agency for Healthcare Research and Quality.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Shachar C, Engel J, Elwyn G. Implications for Telehealth in a Postpandemic Future: Regulatory and Privacy Issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
2. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
3. Rodriguez JA, Clark CR, Bates DW. Digital Health Equity as a Necessity in the 21st Century Cures Act Era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
4. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
5. Lin CC, Dievler A, Robbins C, Sripipatana A, Quinn M, Nair S. Telehealth In Health Centers: Key Adoption Factors, Barriers, And Opportunities. Health Aff (Millwood) 2018 Dec;37(12):1967-1974. [doi: 10.1377/hlthaff.2018.05125] [Medline: 30633683]
6. Fischer SH, Ray KN, Mehrotra A, Bloom EL, Uscher-Pines L. Prevalence and Characteristics of Telehealth Utilization in the United States. JAMA Netw Open 2020 Oct 01;3(10):e2022302 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22302] [Medline: 33104208]
7. Welch BM, Harvey J, O'Connell NS, McElligott JT. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. BMC Health Serv Res 2017 Nov 28;17(1):784. [doi: 10.1186/s12913-017-2744-8] [Medline: 29183372]
8. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization Barriers and Medical Outcomes Commensurate With the Use of Telehealth Among Older Adults: Systematic Review. JMIR Med Inform 2020 Aug 12;8(8):e20359 [FREE Full text] [doi: 10.2196/20359] [Medline: 32784177]
9. Labgold K, Hamid S, Shah S, Gandhi NR, Chamberlain A, Khan F, et al. Measuring the missing: greater racial and ethnic disparities in COVID-19 burden after accounting for missing race/ethnicity data. medRxiv. Preprint posted online October 2, 2020. [doi: 10.1101/2020.09.30.20203315]
10. Long JA, Bamba MI, Ling B, Shea JA. Missing race/ethnicity data in Veterans Health Administration based disparities research: a systematic review. J Health Care Poor Underserved 2006 Feb;17(1):128-140. [doi: 10.1353/hpu.2006.0029] [Medline: 16520522]

Edited by C Basch; submitted 27.08.20; peer-reviewed by P Clardy, C Kruse; comments to author 02.10.20; revised version received 13.10.20; accepted 25.03.21; published 20.05.21.

Please cite as:
Stevens JP, Mechanic O, Markson L, O'Donoghue A, Kimball AB
Telehealth Use by Age and Race at a Single Academic Medical Center During the COVID-19 Pandemic: Retrospective Cohort Study
J Med Internet Res 2021;23(5):e23905
URL: https://www.jmir.org/2021/5/e23905
doi:10.2196/23905
PMID:33974549

©Jennifer P Stevens, Oren Mechanic, Lawrence Markson, Ashley O'Donoghue, Alexa B Kimball. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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

Capturing COVID-19–Like Symptoms at Scale Using Banner Ads on an Online News Platform: Pilot Survey Study

Brian E Dixon[1,2], MPA, PhD; Sumit Mukherjee[3], PhD; Ashley Wiensch[2], MPH; Mary L Gray[4,5], PhD; Juan M Lavista Ferres[3], MSc; Shaun J Grannis[2,6], MD, MSc

[1]Department of Epidemiology, Richard M Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

[2]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

[3]AI for Good Research Lab, Microsoft Corporation, Redmond, WA, United States

[4]New England Lab, Microsoft Research, Cambridge, MA, United States

[5]Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States

[6]Department of Family Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Brian E Dixon, MPA, PhD
Department of Epidemiology
Richard M Fairbanks School of Public Health
Indiana University
1050 Wishard Boulevard, Floors 5 and 6
Indianapolis, IN, 46202
United States
Phone: 1 317 278 3072
Email: bedixon@regenstrief.org

Abstract

Background: Identifying new COVID-19 cases is challenging. Not every suspected case undergoes testing, because testing kits and other equipment are limited in many parts of the world. Yet populations increasingly use the internet to manage both home and work life during the pandemic, giving researchers mediated connections to millions of people sheltering in place.

Objective: The goal of this study was to assess the feasibility of using an online news platform to recruit volunteers willing to report COVID-19–like symptoms and behaviors.

Methods: An online epidemiologic survey captured COVID-19–related symptoms and behaviors from individuals recruited through banner ads offered through Microsoft News. Respondents indicated whether they were experiencing symptoms, whether they received COVID-19 testing, and whether they traveled outside of their local area.

Results: A total of 87,322 respondents completed the survey across a 3-week span at the end of April 2020, with 54.3% of the responses from the United States and 32.0% from Japan. Of the total respondents, 19,631 (22.3%) reported at least one symptom associated with COVID-19. Nearly two-fifths of these respondents (39.1%) reported more than one COVID-19–like symptom. Individuals who reported being tested for COVID-19 were significantly more likely to report symptoms (47.7% vs 21.5%; $P<.001$).

Symptom reporting rates positively correlated with per capita COVID-19 testing rates ($R^2=0.26$; $P<.001$). Respondents were geographically diverse, with all states and most ZIP Codes represented. More than half of the respondents from both countries were older than 50 years of age.

Conclusions: News platforms can be used to quickly recruit study participants, enabling collection of infectious disease symptoms at scale and with populations that are older than those found through social media platforms. Such platforms could enable epidemiologists and researchers to quickly assess trends in emerging infections potentially before at-risk populations present to clinics and hospitals for testing and/or treatment.

(J Med Internet Res 2021;23(5):e24742) doi:10.2196/24742

KEYWORDS

COVID-19; coronavirus; epidemiology; research subject recruitment; signs and symptoms

XSL•FO
RenderX

Introduction

The global outbreak of SARS-CoV-2 led the World Health Organization (WHO) to declare a pandemic on March 11, 2020 [1]. SARS-CoV-2 causes COVID-19, which has a range of manifestations from asymptomatic infection to severe pneumonia, potentially leading to intensive care utilization and death. With over 34 million cases globally, COVID-19 has impacted health and health care in every country. Although COVID-19 spread has leveled off in some parts of the world [2], public health experts anticipate future outbreaks given that only a fraction of the population has been infected with the disease [3].

Given the need to isolate or quarantine infected individuals in order to slow community-level spread, a key component of an effective response to COVID-19 is early identification of new cases. At the beginning of the pandemic, most nations lacked capacity to confirm positive cases, typically performed using a real-time reverse transcription–polymerase chain reaction laboratory test. Over time, public health as well as hospital and commercial labs have expanded their abilities to perform COVID-19 tests. However, COVID-19 is challenging to identify because many cases are mild or asymptomatic. Between 40% and 45% of active viral cases do not exhibit symptoms [3,4]. Other individuals experience more mild symptoms, such as muscle aches, that may not prompt them to seek medical care. Individuals who do not feel sick are less likely to present to hospital or clinic for testing. Therefore, many COVID-19 cases go unreported, making it challenging for epidemiologists to track COVID-19 spread or model its future impact.

There exist several infodemiology [5,6] and digital surveillance tools to identify or estimate suspected COVID-19 cases in a population. These tools seek to capture COVID-19 symptoms, risks, and exposure information directly from consumers. Several recent studies report efforts to capture COVID-19 symptom data from consumers using internet-based surveys [7,8] or social media platforms [9,10]. Additional studies examine the use of mobile apps that allow patients to self-report symptoms; some are dedicated to COVID-19, while others are extensions of existing platforms [11-13]. Further, some studies explore the use of web search behaviors to detect potential clusters of COVID-19 [14].

Existing COVID-19 surveillance tools tend to focus on tracking COVID-19 cases in a localized geographic area. For example, a review of COVID-19 mobile apps found that more than half were official apps managed by local health authorities presumably focused in their local region [13]. Moreover, many reports on existing tools either lack details regarding users who provided information [12,14] or the population surveyed was skewed, such as in Shen et al [9] where 85% of respondents were between 19 and 40 years of age. To maximize efficacy, digital public health surveillance tools should capture data from a representative sample across local, state, and national levels to enable global outbreak identification and trend monitoring. Furthermore, public health authorities should use a variety of surveillance tools to gather data from multiple sources to triangulate disease outbreaks and trends.

In this study, we sought to examine the feasibility of using an existing web-based news platform to recruit individuals willing to self-report symptoms associated with COVID-19 as well as health behaviors via a survey. We sought to recruit a diverse set of volunteers in a manner that can be scaled during a pandemic.

Methods

Overview

We employed an epidemiologic survey to collect COVID-19 symptoms and behaviors from individuals recruited using a web-based news platform. Not knowing whether the method would be successful, we approached the project as a feasibility study. We collected data in multiple countries over 4 weeks to examine the platform's ability to capture a representative sample of the underlying population with internet access in order to detect suspected COVID-19 cases before individuals may have sought diagnosis and treatment from the health system.

Survey Instrument

The survey primarily focused on capturing symptoms associated with COVID-19. Using the best available data at the time (early April 2020), the following symptoms were included in the survey: fever, cough, itchy or watery eyes, loss of sense of taste or smell, nasal congestion or runny nose, sore throat, and shortness of breath or difficulty breathing. Our list was principally based on the list published by the US Centers for Disease Control and Prevention, which recognized three main symptoms for COVID-19: fever, cough, and shortness of breath [15]. An early systematic review, published in March 2020, that included 1576 early COVID-19 patients reported that the most prevalent clinical symptom was fever, followed by cough, fatigue, and dyspnea [16]. A later review reported the main symptoms to be fever, cough, fatigue, slight dyspnea, sore throat, headache, conjunctivitis, and gastrointestinal issues [17]. We further included anosmia (loss of the sense of smell) and ageusia (loss of the sense of taste), because early evidence [18] linking these symptoms was emerging from the literature around the time of survey development. Because these symptoms were only associated with COVID-19, and they are similar to symptoms experienced by patients with other diseases, we considered these symptoms to indicate COVID-19–like illness rather than a definitive case of COVID-19, which is confirmed through laboratory testing.

In addition to symptoms, we asked respondents to report some behaviors linked to COVID-19 risk. We specifically asked respondents to describe where they had spent time outside of their home in the past 2 weeks, more than 18 miles from home or fewer than 18 miles from home. We also asked if they had traveled outside of their home country in the previous 2 weeks.

We collected minimal demographics for respondents to protect privacy. The only demographics requested were age group (eg, 40-49 years) and gender. We further asked for respondents' postal codes (eg, ZIP Code) to enable comparisons with localized outbreaks of COVID-19. We did not ask respondents to separately report their country of residence. We did not

XSL•FO
RenderX

capture race data as it is not customary to collect such information outside the United States.

We created the survey in English and subsequently translated it into 12 additional languages for release into 30 different language markets. Languages included, among others, Mandarin, French, German, Japanese, Greek, and Spanish.

Subject Recruitment

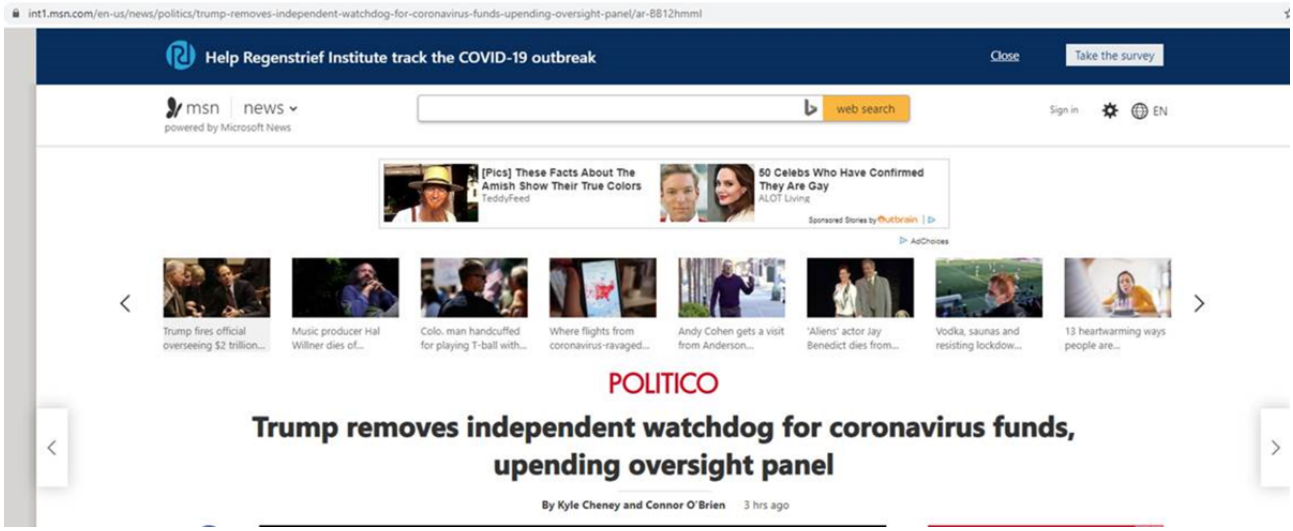
To recruit participants, we placed banner ads on the Microsoft News platform. Microsoft News delivers news from multiple, popular publishers across web and mobile experiences. The service is used by nearly a half billion monthly users, making it one of the largest audiences of news readers in the world [19].

The Microsoft News team embedded a link to the Regenstrief Institute survey on its news sites in 31 global language markets to reach a large, general population sample for the survey. Each language market directed users to a survey in that language. Initially, Microsoft News editors manually embedded a promotion button calling for survey participants that linked to the survey. Editors randomly embedded the link in approximately 150 pieces of news content daily that were delivered to readers across the United States. Editors placed a button on news content at randomized time windows (eg, 5:36 AM Pacific Time [PT]; 2:12 PM PT) to limit the time window in an effort to deter malicious tampering (eg, bot attacks) during a 1-week survey pilot period. Unique URLs for each language-specific survey were also used to prevent bot attacks. Editors also attempted to avoid placing the button on COVID-19–related content to prevent oversampling people reading about COVID-19. Moreover, editors tried to ensure the button was placed at a low profile (ie, "below the fold" of the news article).

By April 10, 2020, COVID-19–related content accounted for 80% of US news traffic. This challenged the manual embedding process, placing an unmanageable burden on the Microsoft News editorial team. Therefore, in consultation with the research team, Microsoft News switched to a conditional banner method on April 11, 2020.

The conditional banner ad methodology involved automatically rendering the ad (Figure 1) at the top of all news articles, but with the condition that a user would only see the banner twice in 30 days. The conditional banner provided a gating effect so that Microsoft News readers could only be served a total of two conditional banners calling for participants to complete the survey during any single month. For example, if a reader visited a travel article and a politics article during the month, the reader had two opportunities to see the banner ad but would not see it again until the next month. Furthermore, once a reader clicked on a conditional banner ad, which took them to the survey, they would not receive another conditional banner ad calling for participants until the next month. We offered users two banner ads per month until they completed the survey or until the survey closed. Although Microsoft News tracked whether a user saw the banner ad more than once, user identifiers were not provided to the research team, so we could not identify individuals or link their multiple responses.

Figure 1. Screenshot of the Microsoft News platform showing the banner ad placed at the top of the screen that recruited participants to take a survey about COVID-19 symptoms and behaviors.



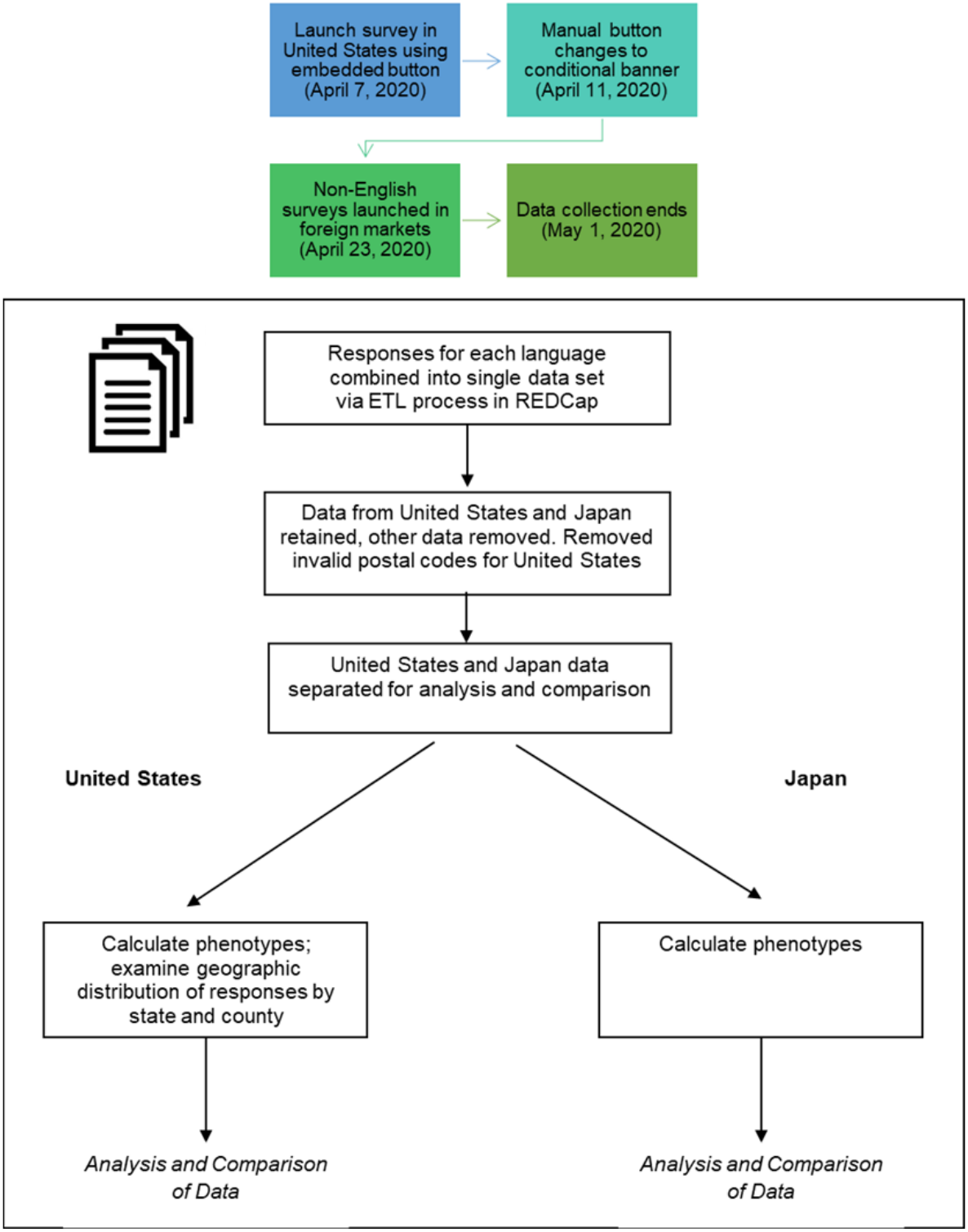
Importantly, Microsoft News readers who clicked on the banner ad always saw an initial screen on the survey site that prompted them for their explicit consent to participate, informed them of the risks of participating, and gave them a chance to leave the survey and either return to Microsoft News or access COVID-19 information resources (eg, WHO site). No user data were captured until participants consented. The Regenstrief Institute monitored referring websites to confirm that nearly all respondents came to the survey from Microsoft News.

Data Collection

Surveys were developed and rolled out in two phases. We first generated and published the US-based English survey. The US-based English survey was published beginning on April 7, 2020. Beginning April 11, 2020, the conditional banner ad replaced the manually embedded button. Non-English surveys were published on April 23, 2020. All surveys closed on May 1, 2020. Figure 2 depicts the overall flow for collection and analysis of data.

XSL•FO
RenderX

Figure 2. Overview of data collection and analysis processes. ETL: extract, transform, load; REDCap: Research Electronic Data Capture.



Survey responses were collected and managed in collaboration with Research Technologies, a Pervasive Technology Institute Center at Indiana University (IU) [20], using Research Electronic Data Capture (REDCap) tools [21,22]. Data from each language-based survey were stored in separate tables on the REDCap server, then extracted into a combined data set for analysis using REDCap-ETL (extract, transform, load) and the IU REDCap-ETL Gateway [23,24]. The language of the respondent was captured as a distinct field in the combined data set.

REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources. REDCap-ETL extracts

XSL•FO
RenderX

data from REDCap, transforms the extracted data, and loads the transformed data into a database. The IU REDCap-ETL Gateway enables high-throughput data transfer between REDCap and multiple endpoints by allowing REDCap-ETL to scale to many simultaneous runs using the job scheduler of a high-performance computer.

Analytical Methods

We used descriptive statistics to summarize responses overall and by country to examine patterns. We further employed ecological methods to compare prevalence by country as well as association methods to examine differences in responses by country. Moreover, we employed correlation methods to compare symptom responses to standard outbreak measures. More than 85% of responses were received from residents in the United States and Japan. Therefore, we only included these responses in the analysis.

We first examined patterns regarding the limited demographics captured in the survey. Descriptive statistics provide a sense for the representativeness of the respondents. We calculated summary statistics for respondent demographics after removing all entries that did not provide informed consent, indicated by failing to click *Next*. In all demographic categories, we restricted our summarization to those individuals who did not select *prefer not to say*.

We next summarized responses based on respondents' reported symptoms and behaviors that might contribute to COVID-19 exposure. We asked respondents to indicate the presence of one or more current symptoms, including symptoms believed to be associated with COVID-19 infection and those not shown to be associated. The symptoms believed to be more common among patients testing positive for COVID-19 include fever, cough, shortness of breath, loss of sense of taste or smell, and sore throat [7,17,18,25].

Symptoms associated with COVID-19 were grouped into various phenotypes for COVID-19–like illness. We defined four different COVID-19 phenotypes for analysis, summarized in Table 1. We examined a variety of phenotypes to explore symptom combinations associated with COVID-19. We examined multiple phenotypes since evidence suggests that many COVID-19 patients experience more than one symptom, but a specific set of symptoms is not yet definitive. Phenotype 3 is the broadest definition for a potential case, as individuals could report any of the symptoms previously known to be associated with COVID-19 infection. Phenotype 4 is the most restrictive phenotype, as it requires indication of fever or loss of taste or smell plus one of the other symptoms associated with infection. All phenotypes were informed by available evidence [26] indicating which symptoms or combinations are associated with infection.

Table 1. Phenotypes for suspected COVID-19 cases.

Symptoms	Phenotype 1	Phenotype 2 (one or the other)		Phenotype 3	Phenotype 4 (both must apply)	
	Individual reports *at least* 2 of the following	Individual reports *any* 1 of the following	Individual reports *at least* 2 of the following	Individual reports *any* 1 of the following	Individual reports *at least* 1 of the following	Individual reports *any* 1 of the following
Fever	✓		✓	✓	✓	
Cough	✓	✓		✓		✓
Loss of sense of taste or smell	✓		✓	✓	✓	
Sore throat	✓		✓	✓		✓
Shortness of breath or difficulty breathing	✓	✓		✓		✓

To visualize the distribution of respondents in the United States, we restricted the data to valid US ZIP Codes. The text field used by respondents to enter their postal code was unrestricted, given the lack of global conventions for representing residential area. Invalid ZIP Code entries were ignored when generating maps of US respondents. States were assigned based on published ZIP Code values associated with each state. To map ZIP Code values to US counties, we leveraged a crosswalk between county Federal Information Processing System codes and ZIP Code.

Chi-square tests of independence were used to examine whether responses differed significantly between the United States and Japan. We further compared responses between respondents indicating that they received testing versus those reporting that they did not receive laboratory testing for COVID-19. We employed correlation to compare the proportion of respondents meeting various phenotype definitions in a given state to existing surveillance per capita metrics reported in the same state during the study period. First, we correlated phenotypes with the per capita testing rate in each state, since individuals with symptoms are theoretically more likely to seek out testing. Second, we correlated phenotypes with the reported per capita case rate, which represents the positive case volume in a given state. These metrics were obtained from The Atlantic application programming interface [27], which are aggregated from a variety of sources.

Ethics Approval

This study was approved by the Institutional Review Board at IU as exempt research.

Results

We received a total of 87,322 valid responses. Of these, 47,424 (54.3%) were from the United States and 27,936 (32.0%) were from Japan. Table 2 summarizes responses and the demographics of respondents. Responses from Japan are skewed toward male gender (21,220/27,936, 77.0%), whereas the US

XSL•FO
RenderX

responses are more balanced (male: 25,787/47,424, 57.0%). With respect to age, more than half of the respondents from both countries were over 50 years of age. Chi-square analysis found the rates of responses in Japan were significantly different from those in the United States.

The distribution of respondents in the United States is depicted in Figure 3 as well as in Multimedia Appendix 1. The raw counts of respondents from each state and ZIP Code (Figures S1 and S2 in Multimedia Appendix 1) show broad geographical distribution of responses across the nation. Higher numbers of responses were generally observed in areas with higher populations. When normalized by state population, responses were greatest in the northern states, especially the extreme northeastern states and the states in the Upper Plains region.

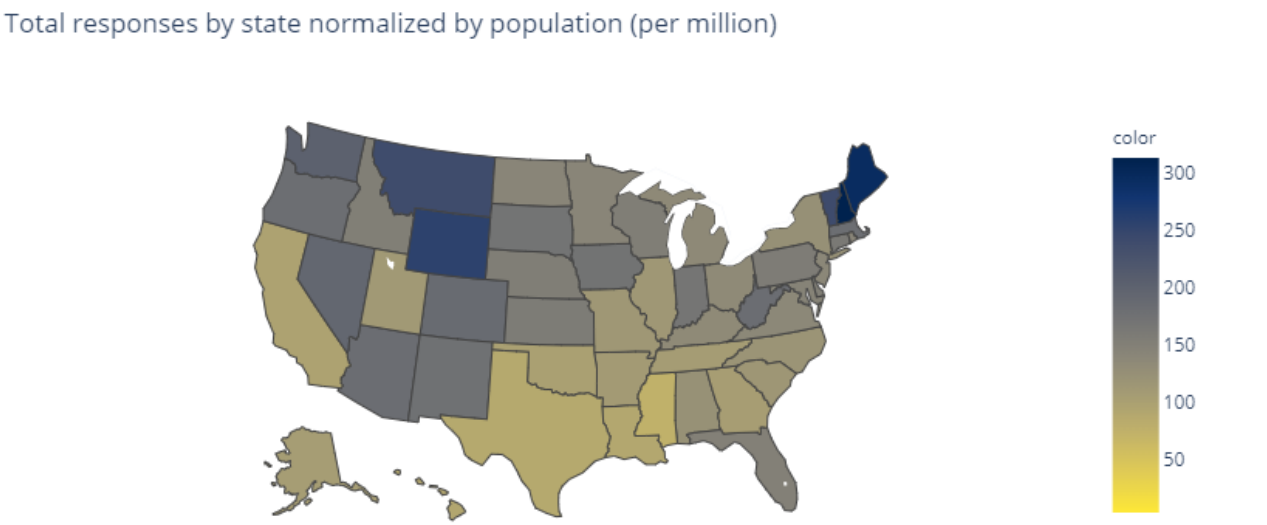
Table 2. Demographics of all respondents (who completed the consent portion) of a web-based questionnaire about COVID-19 symptoms recruited through the Microsoft News platform.

Demographic	All respondents, n (%)[a]	Respondents from the United States, n (%)[a]	Respondents from Japan, n (%)[a]	*P* value[b]
Age (years) (all: n=83,901; United States: n=45,050; Japan: n=27,536)				
80+	2672 (3.2)	2202 (4.9)	252 (0.9)	<.001
70-79	12,150 (14.5)	8402 (18.7)	2343 (8.5)	<.001
60-69	23,632 (28.2)	14,438 (32.0)	6233 (22.6)	<.001
50-59	23,174 (27.6)	10,635 (23.6)	9539 (34.6)	<.001
40-49	12,580 (15.0)	5040 (11.2)	5850 (21.2)	<.001
30-39	5208 (6.2)	2567 (5.7)	1757 (6.4)	<.001
18-29	4485 (5.3)	1766 (3.9)	1562 (5.7)	<.001
Gender (all: n=84,134; United States: n=45,195; Japan: n=27,461)				
Female	29,171 (34.7)	19,221 (42.5)	6174 (22.5)	<.001
Male	54,651 (65.0)	25,787 (57.1)	21,220 (77.3)	<.001
Transgender	180 (0.2)	102 (0.2)	40 (0.1)	<.001
Nonbinary	132 (0.2)	85 (0.2)	27 (0.1)	<.001

[a]Percentages may not add up to 100% due to rounding.

[b]*P* values were based on chi-square analysis.

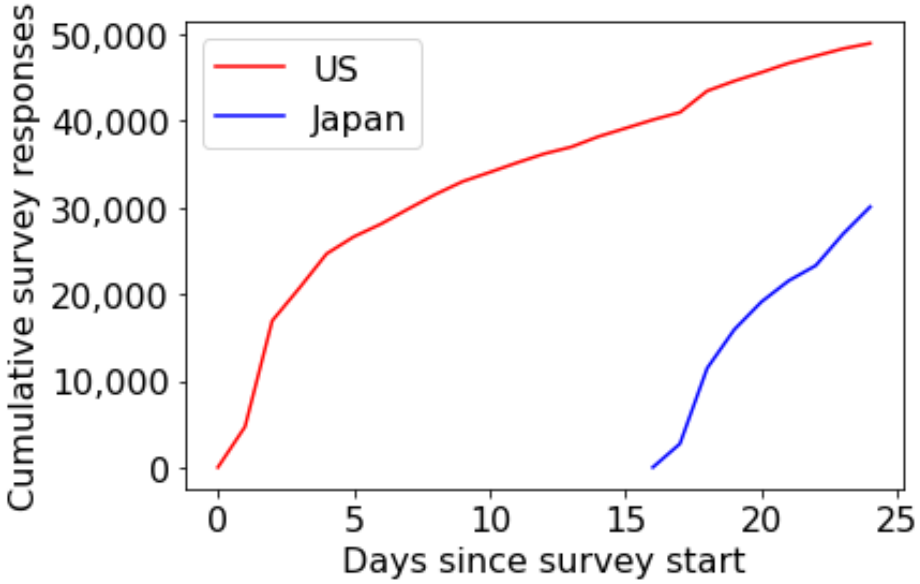
Figure 3. Total responses by US state normalized by population. Lighter colors indicate low counts, whereas darker colors indicate a higher number of responses.



XSL•FO
RenderX

Responses over time are summarized in Figure 4, stratified by the location of the response. Cumulative responses increased rapidly in the days immediately after the survey launch followed by a continued, steady increase in the weeks after the launch. The Japanese survey was launched 16 days after the English survey, yet responses followed a similar pattern with respect to rapid increase following the launch.

Figure 4. Cumulative responses over time, stratified by location of respondent, to a survey on COVID-19 symptoms. The survey was offered in Japan 16 days after it was introduced in the United States.



In Table 3 we summarize the responses with respect to COVID-19 symptoms, phenotypes, and behaviors. Most respondents (56,551/87,322, 64.8%) reported no symptoms. Among those reporting at least one symptom (30,771/87,322, 35.2%), a total of 19,631 respondents (63.8%) reported at least one of the following symptoms associated with COVID-19: fever, cough, loss of sense of taste or smell, sore throat, shortness of breath, or difficulty breathing. These respondents met the inclusion criteria for Phenotype 3. Higher proportions of respondents in the United States reported COVID-19–like symptoms than respondents in Japan across all symptoms and phenotypes (42.2% vs 25.2%; $P<.001$).

Among the 19,631 respondents meeting criteria for Phenotype 3, 7670 (39.1%) reported at least two symptoms associated with COVID-19 (Phenotype 1). A total of 15,341 out of 30,771 individuals with symptoms (49.9%) reported either cough or shortness of breath, or they reported at least two symptoms from the set of fever, sore throat, or loss of taste or smell (Phenotype 2). Examining the more specific symptoms of fever and loss of sense of taste or smell, 4820 out of 30,771 individuals with symptoms (15.7%) reported at least one of these symptoms in combination with one of the other symptoms (Phenotype 4). Reported symptoms met phenotype definitions at higher rates among respondents from the United States compared to respondents from Japan ($P<.001$).

Compared to respondents who did not report symptoms, those reporting at least one of the symptoms associated with COVID-19 (Phenotype 3) in the United States were more likely to report getting tested ($P<.001$). Similar rates were observed in Japan. Overall, higher proportions of respondents in the United States reported getting tested than did those in Japan ($P<.001$). In both countries, individuals meeting the inclusion criteria for one of the phenotypes more frequently reported that they had traveled more than 30 km from their home within the past 14 days. Chi-square analysis found that the rates of responses in Japan were significantly different from those in the United States.

XSL•FO
RenderX

Table 3. Symptoms, phenotypes, and behaviors reported by respondents using a web-based questionnaire recruited through the Microsoft News platform.

Symptoms, phenotypes, and behaviors	All respondents (N=87,322), n (%)	Respondents in the United States (n=47,424), n (%)	Respondents in Japan (n=27,936), n (%)	*P* value[a]
Individual symptom				
Fever	3308 (3.8)	2428 (5.1)	527 (1.9)	<.001
Loss of taste or smell	2585 (3.0)	2032 (4.3)	214 (0.8)	<.001
Cough	12,826 (14.7)	9189 (19.4)	2184 (7.8)	<.001
Shortness of breath	5768 (6.6)	4502 (9.5)	495 (1.8)	<.001
Sore throat	8021 (9.2)	5135 (10.8)	1873 (6.7)	<.001
Runny nose	17,157 (19.6)	11,772 (24.8)	3567 (12.8)	<.001
Itchy or watery eyes	12,800 (14.7)	8866 (18.7)	2518 (9.0)	<.001
None	56,551 (64.8)	27,411 (57.8)	20,888 (74.8)	<.001
At least 1 symptom	30,771 (35.2)	20,013 (42.2)	7048 (25.2)	<.001
Phenotype				
Phenotype 1	7670 (8.7)	5837 (12.2)	924 (3.3)	<.001
Phenotype 2	15,341 (17.4)	10,961 (23.0)	2498 (8.9)	<.001
Phenotype 3	19,631 (22.3)	13,308 (27.9)	3881 (13.8)	<.001
Phenotype 4	4820 (5.5)	3647 (7.6)	623 (2.2)	<.001
Tested				
Phenotype 1	635 (8.3)	551 (9.4)	23 (2.5)	<.001
Phenotype 2	827 (5.4)	707 (6.5)	32 (1.3)	<.001
Phenotype 3	939 (4.8)	793 (6.0)	45 (1.2)	<.001
Phenotype 4	523 (10.9)	461 (12.6)	21 (3.4)	<.001
None	869 (1.5)	496 (1.8)	227 (1.1)	<.001
Traveled more than 30 km				
Phenotype 1	1539 (20.1)	1242 (21.3)	164 (17.7)	<.001
Phenotype 2	2724 (17.8)	2052 (18.7)	436 (17.5)	<.001
Phenotype 3	3394 (17.3)	2433 (18.3)	659 (17.0)	<.001
Phenotype 4	994 (20.6)	799 (21.9)	107 (17.2)	<.001
None	9237 (16.3)	5070 (18.5)	3181 (15.2)	<.001

[a]*P* values were based on chi-square analysis.

In Table 4, we compare respondents who reported that they were tested for COVID-19 versus those who reported no testing. Overall, few respondents (1967/87,322, 2.3%) indicated that they were tested for COVID-19. Examining symptom combinations and phenotypes revealed that individuals who reported symptoms were more likely to report being tested for COVID-19. Among those tested for COVID-19, nearly half (939/1967, 47.7%) reported at least one symptom compared to individuals who were not tested (18,197/84,498, 21.5%; $P<.001$). The most prevalent symptom combinations among individuals tested for COVID-19 included cough and nasal congestion or runny nose (404/1967, 20.5%), cough and shortness of breath or difficulty breathing (372/1967, 18.9%), as well as fever and cough (318/1967, 16.2%). Nearly two-thirds (55,084/84,498, 65.2%) of individuals who reported no symptoms also reported not being tested.

The proportion of respondents in each state meeting a given phenotype definition positively correlated with the proportion of individuals tested in that state (Figures S1-S4 in Multimedia Appendix 2). For example, the proportion of individuals who reported symptoms corresponding to Phenotype 3 were significantly ($P<.001$) positively ($R^2=0.26$) correlated with the per capita rate of individuals tested for COVID-19 (Figure S3 in Multimedia Appendix 2). The proportion of respondents meeting a given phenotype definition showed slight correlation with the reported case rate ($R^2\leq0.04$) (Figures S5-S8 in Multimedia Appendix 2), yet these correlations were not statistically significant ($P\geq.05$). For example, the proportion of individuals who reported symptoms corresponding to Phenotype 4 were nonsignificantly ($P=.23$) positively ($R^2=0.04$) correlated with the per capita rate of individuals testing positive for COVID-19 (Figure S8 in Multimedia Appendix 2).

XSL•FO
RenderX

Table 4. Symptom combinations and phenotypes among respondents, stratified by whether or not they reported getting tested for COVID-19.

Symptom combinations and phenotypes	All respondents (N=87,322), n (%)	Respondents tested for COVID-19 (n=1967), n (%)	Respondents not tested for COVID-19 (n=84,498), n (%)	*P* value[a]
Individual symptom combination				
Fever and cough	2268 (2.4)	318 (16.2)	1895 (2.2)	<.001
Fever and itchy or watery eyes	1345 (1.4)	146 (7.4)	1162 (1.4)	<.001
Fever and loss of sense of taste or smell	1073 (1.1)	189 (9.6)	850 (1.0)	<.001
Fever and nasal congestion or runny nose	1821 (1.9)	227 (11.5)	1548 (1.8)	<.001
Fever and sore throat	1666 (1.8)	205 (10.4)	1411 (1.7)	<.001
Fever and shortness of breath or difficulty breathing	1626 (1.7)	254 (12.9)	1327 (1.6)	<.001
Cough and itchy or watery eyes	4713 (5.0)	246 (12.5)	4338 (5.1)	<.001
Cough and loss of sense of taste or smell	1625 (1.7)	244 (12.4)	1327 (1.6)	<.001
Cough and nasal congestion or runny nose	6523 (6.9)	404 (20.5)	5963 (7.1)	<.001
Cough and sore throat	4056 (4.3)	310 (15.8)	3654 (4.3)	<.001
Cough and shortness of breath or difficulty breathing	3650 (3.9)	372 (18.9)	3186 (3.8)	<.001
Itchy or watery eyes and loss of sense of taste or smell	1228 (1.3)	129 (6.6)	1056 (1.2)	<.001
Itchy or watery eyes and nasal congestion or runny nose	7382 (7.8)	271 (13.8)	6932 (8.2)	<.001
Itchy or watery eyes and sore throat	3166 (3.4)	173 (8.8)	2921 (3.5)	<.001
Itchy or watery eyes and shortness of breath or difficulty breathing	2430 (2.6)	173 (8.8)	2179 (2.6)	<.001
Loss of sense of taste or smell and nasal congestion or runny nose	1585 (1.7)	209 (10.6)	1324 (1.6)	<.001
Loss of sense of taste or smell and sore throat	1147 (1.2)	162 (8.2)	945 (1.1)	<.001
Loss of sense of taste or smell and shortness of breath or difficulty breathing	1241 (1.3)	194 (9.9)	1000 (1.2)	<.001
Nasal congestion or runny nose and sore throat	4079 (4.3)	250 (12.7)	3729 (4.4)	<.001
Nasal congestion or runny nose and shortness of breath or difficulty breathing	3125 (3.3)	272 (13.8)	2765 (3.3)	<.001
Sore throat and shortness of breath or difficulty breathing	2112 (2.2)	226 (11.5)	1826 (2.2)	<.001
Phenotype				
Phenotype 1	7670 (8.1)	635 (32.3)	6857 (8.1)	<.001
Phenotype 2	15,341 (16.3)	827 (42.0)	14,117 (16.7)	<.001
Phenotype 3	19,631 (20.8)	939 (47.7)	18,197 (21.5)	<.001
Phenotype 4	4820 (5.1)	523 (26.6)	4166 (4.9)	<.001
No symptoms	56,551 (60.0)	869 (44.2)	55,084 (65.2)	<.001

[a]*P* values were based on chi-square analysis.

XSL•FO
RenderX

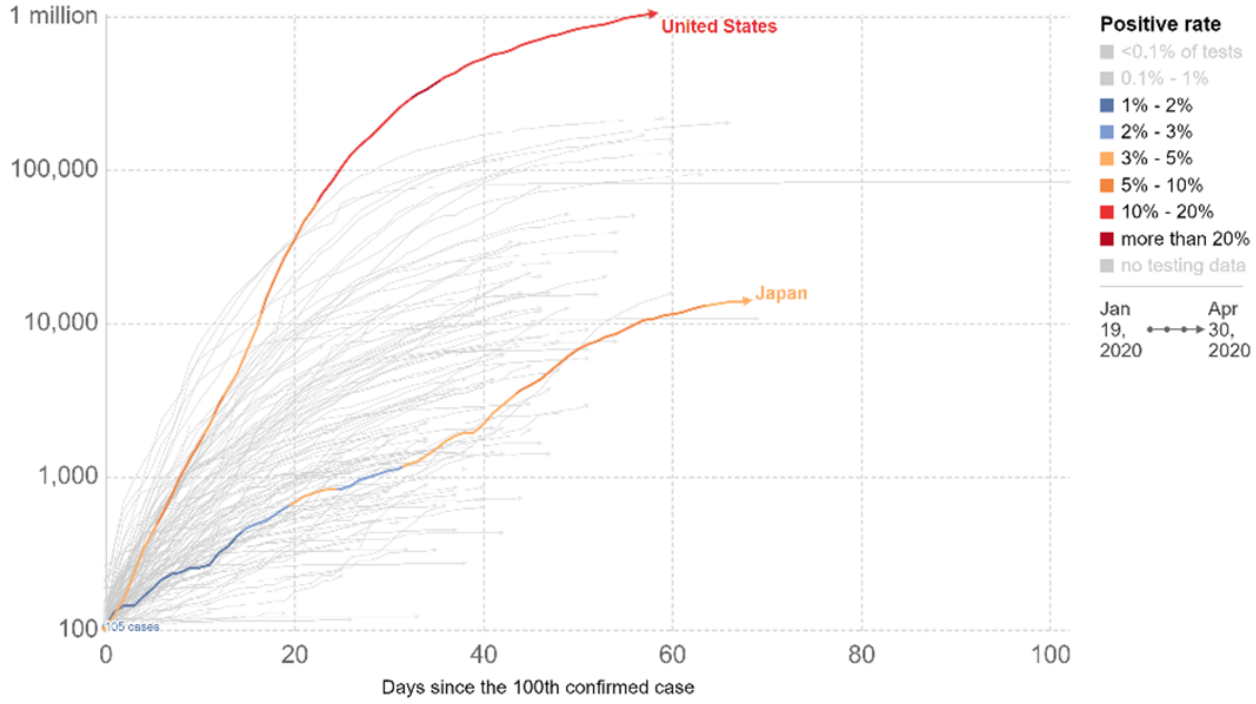
Discussion

Principal Findings

In this study, we recruited tens of thousands of respondents within 3 weeks to an online survey asking questions about symptoms associated with a novel infectious disease quickly spreading across the globe using banner ads on an online news platform. A principal finding is that an internet news platform is feasible for quickly collecting large-scale data from individuals across local, state, and national areas. Furthermore, this recruitment method reached an older population at greater risk of complications from COVID-19, distinguishing our study from prior research. Whereas many infodemiology studies recruit subjects from social media platforms where users are primarily under 40 years of age, the news platform we leveraged draws primarily from populations at least 50 years of age. Both findings have implications for future surveillance efforts in public health.

At time of the survey, Japan's epidemic curve had started to flatten, with daily increases in new cases between 3% and 5%, whereas the curve in the United States continued its upward trajectory, with daily increases between 10% and 20% (Figure 5 [26,28]). On April 30, 2020, the United States reported a daily rate of new infections of 8.26 per 100,000 population. On the same date, Japan reported a daily rate of 0.19 new infections per 100,000 population. In this study, rates of reported symptoms in Japan were significantly lower than in the United States, providing face validity for respondents' self-reported data. This further builds the case for an online news platform as a reliable method for recruiting survey respondents.

Figure 5. COVID-19 epidemic curves for the United States and Japan through May 1, 2020 [26]. The plot shows cumulative confirmed COVID-19 cases. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing. The source of this data is the European Centre for Disease Prevention and Control, Situation Update Worldwide, last updated August 31, 2020, 10:34 AM (London time). Official data were collated by Our World in Data [28].



The survey data further yielded interesting insights about the pattern of symptoms reported by respondents. First, nearly two-thirds of those reporting at least one symptom reported at least two symptoms. Emerging evidence suggests that symptom combinations may be predictive of COVID-19 infection [26]. The fact that most individuals reporting COVID-19–like symptoms reported more than one symptom provides further evidence that exploring specific phenotypes might be fruitful for future research. Second, the constellation of symptoms involving loss of taste or smell with fever appears more often among individuals who reported that they received testing for COVID-19. Several studies report loss of taste and smell as common among COVID-19 patients [18,26,29], suggesting that these symptoms cause individuals to seek out medical care or indicate an association with infection. Moreover, those reporting at least one symptom were significantly more likely to report having been tested for COVID-19 than those without symptoms, and those reporting symptoms were significantly associated with testing rates. Yet the rates of individuals reporting symptoms were not significantly associated with positive case rates, which were admittedly hindered due to limited testing infrastructure in the United States at the time of the survey [30]. These findings are novel, given the data were captured from community residents rather than individuals hospitalized from COVID-19, which is the population most prior studies used to report common symptoms reported by individuals infected with SARS-CoV-2.

Implications of Findings

Based on these findings, we encourage further study of online news platforms as a tool for reaching broad and diverse populations by public health agencies during emerging outbreaks. Although the news platform was male dominated

XSL•FO
RenderX

and skewed older with respect to population, we recommend its use in combination with other epidemiologic methods, including infodemiology, random selection, and contact tracing. Only by using multiple methods will public health organizations successfully achieve diverse representation in the total set of data used to inform disease tracking efforts. Surveying only young, healthy populations is not sufficient for ascertaining prevalence of diseases like COVID-19. Yet inclusion of younger population data is important to understand disease spread during a crisis. It is possible that other news platforms may have younger audiences or cater to specific audiences that would improve diversity. Capturing data across platforms and methods to integrate findings would benefit future infodemiology research.

The methods used in this study go beyond the typical methods used in public health to ascertain symptoms among individuals who present at clinical or public health sites. Relying on data from individuals who present physically for treatment or testing alone is not sufficient to generate population-level information about the epidemiology of an emerging infectious disease, as only some individuals present for care or testing. Recall that approximately 40% of individuals with COVID-19 are asymptomatic [3], and many experience mild symptoms that do not require clinical intervention. Moreover, we received responses from individuals in small towns and rural communities, areas that are often medically underserved, thereby limiting patients' ability to present for care. Furthermore, rural areas often lack robust laboratory testing infrastructures, a broad challenge that COVID-19 illuminated. Finally, the methods used here allowed for at-scale collection of data in a way that respects individuals' privacy, as respondents did not need to provide any data that might otherwise be used to link back to medical records or sensitive information.

Internet-based survey methods are needed to gather data quickly during a crisis to inform where the disease might be present or where it might be headed next. Modeling efforts at the start of the COVID-19 pandemic were limited by a paucity of input data as well as a priori assumptions that drive model parameters. Having intelligence on symptoms from a broad population as well as behaviors (eg, those travelling >30 km from home) could better inform model development. Moreover, understanding which populations have received testing, and which ones do not appear to have access to testing, can inform deployment of test kits by national and/or state public health agencies. Finally, identification of symptom hot spots could help drive deployment of personal protective equipment as well as health care workers.

Limitations and Future Directions

Despite the promise of the methods described in this study, there remain many challenges for using internet news platforms and other infodemiology methods to inform public health action. Given limited testing infrastructure in many states, analysis of survey data in comparison with later waves of the pandemic might better show how survey data correlate with standard epidemiologic metrics, such as testing rates, case rates, and mortality. Furthermore, the descriptive nature of this study prevented examination of potential sampling bias in respondents. We observed some skewing related to gender and age. Further exploration of the methods for ad placement would be necessary to establish news platforms as a reliable method for infodemiology.

The study was further limited by the location data being restricted to postal codes. Postal code structures are not standardized globally. Some postal codes repeat in the world, which prevented accurate mapping of some responses to a correct country or province. Furthermore, some entries in the postal code field were unintelligible. It is also possible that some of the Japanese responses occurred outside Japan, as postal code information was self-reported rather than verified with location services within the web browser or mobile device. Moreover, the survey did not ask whether those tested for SARS-CoV-2 virus tested positive or negative. Therefore, we considered individuals reporting symptoms to have a COVID-19–like illness, recognizing that COVID-19 symptoms are similar to several respiratory diseases, such as influenza. The symptom list was further restricted to the best available evidence at the time of survey development.

In the future, we plan to expand our work to extend the survey to additional audiences and in combination with other methods for reaching broader audiences. We further plan to update the symptom list and capture additional details about an individual and their health history that could be correlated with COVID-19 cases and/or known risk factors for infection as well as hospitalization and death. In addition, because severely ill patients may not be able to complete an online survey, we will also explore whether our survey might be adapted to allow respondents to report information on family members who may be experiencing COVID-19–like symptoms.

Conclusions

News platforms can be used to quickly recruit study participants, enabling collection of infectious disease symptoms at scale and with populations that are distinct from those found through social media platforms. Such platforms could enable epidemiologists and researchers to quickly assess trends in emerging infections potentially before individuals present to clinics and hospitals for testing and/or treatment.

Acknowledgments

The authors would like to thank the REDCap support team, part of Research Technologies at IU, especially Andrew Arenson, for their assistance with configuring our REDCap surveys as well as their support with data management. We further acknowledge the Regenstrief Institute, Inc for its support of this nonfunded research project during the pandemic and the assistance provided by Megan Hall and Cindi Hart, RN, in creating the REDCap surveys. From the Microsoft team, we thank Jeri Condit, Dr Ahmed

XSL•FO
RenderX

Elshafei, and Carmen O'Toole for their support in developing and implementing the survey. We also appreciate the support of Fred Cate, Vice President of Research at IU.

Authors' Contributions

The study was conceived by authors BD, MG, and SG. The coauthors SM, JF, and AW contributed significantly to the acquisition, analysis, and interpretation of data as well as the rigor of the methodology and analysis. The primary author BD drafted the manuscript, and all coauthors significantly contributed to revising the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

MLG, SM, and JMLF are employees of the Microsoft Corporation. The other authors have no conflicts to declare.

Multimedia Appendix 1
Distribution of responses by US state and county.
[DOCX File , 300 KB - jmir_v23i5e24742_app1.docx]

Multimedia Appendix 2
Correlations of individuals meeting phenotype definitions with existing surveillance metrics.
[DOCX File , 132 KB - jmir_v23i5e24742_app2.docx]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-04-29]
2. Adeline S, Hanzhang JC, Hurt A, Wilburn T, Wood D, Talbot R. Tracking coronavirus around the US: See how your state is doing. NPR. URL: https://www.npr.org/sections/health-shots/2020/03/16/816707182/map-tracking-the-spread-of-the-coronavirus-in-the-u-s [accessed 2021-04-29]
3. Menachemi N, Yiannoutsos CT, Dixon BE, Duszynski TJ, Fadel WF, Wools-Kaloustian KK, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample - Indiana, April 25-29, 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 24;69(29):960-964 [FREE Full text] [doi: 10.15585/mmwr.mm6929e1] [Medline: 32701938]
4. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, Imperial College COVID-19 Response Team, Imperial College COVID-19 Response Team. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature 2020 Aug;584(7821):425-429. [doi: 10.1038/s41586-020-2488-1] [Medline: 32604404]
5. 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]
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 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
7. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological observations on the association between anosmia and COVID-19 infection: Analysis of data from a self-assessment web application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]
8. Luo H, Lie Y, Prinzen FW. Surveillance of COVID-19 in the general population using an online questionnaire: Report from 18,161 respondents in China. JMIR Public Health Surveill 2020 Apr 27;6(2):e18576 [FREE Full text] [doi: 10.2196/18576] [Medline: 32319956]
9. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: Observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
10. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: Retrospective big data infoveillance study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
11. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to support COVID-19 education, self-assessment, and symptom monitoring in the Netherlands: Observational study. JMIR Mhealth Uhealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]
12. 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 [FREE Full text] [doi: 10.2196/19902] [Medline: 32568728]

XSL•FO
RenderX

13. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: Content analysis and review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
14. Rovetta A, Bhagavathula AS. COVID-19-related web search behaviors and infodemic attitudes in Italy: Infodemiological study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
15. Symptoms of COVID-19. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [accessed 2021-04-29]
16. 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]
17. Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: A comprehensive review. J Intern Med 2020 Aug;288(2):192-206 [FREE Full text] [doi: 10.1111/joim.13091] [Medline: 32348588]
18. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): A multicenter European study. Eur Arch Otorhinolaryngol 2020 Aug;277(8):2251-2261 [FREE Full text] [doi: 10.1007/s00405-020-05965-1] [Medline: 32253535]
19. Microsoft News. URL: https://www.microsoft.com/en-us/microsoftnews [accessed 2021-04-29]
20. Pervasive Technology Institute, Indiana University. URL: https://pti.iu.edu/ [accessed 2021-04-29]
21. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
22. 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. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Arenson A, Mullen J. The REDCap-ETL external module. In: Proceedings of the 2019 REDCap Conference. 2019 Presented at: 2019 REDCap Conference; September 16-19, 2019; Vancouver, BC URL: http://hdl.handle.net/2022/24650
24. Arenson A, Mullen J. REDCap-ETL Gateway: Using a supercomputer to run REDCap-ETL. In: Proceedings of the 2019 REDCap Conference. 2019 Presented at: 2019 REDCap Conference; September 16-19, 2019; Vancouver, BC URL: http://hdl.handle.net/2022/24649
25. Chow EJ, Schwartz NG, Tobolowsky FA, Zacks RLT, Huntington-Frazier M, Reddy SC, et al. Symptom screening at illness onset of health care personnel with SARS-CoV-2 infection in King County, Washington. JAMA 2020 May 26;323(20):2087-2089 [FREE Full text] [doi: 10.1001/jama.2020.6637] [Medline: 32301962]
26. Dixon BE, Wools-Kaloustian K, Fadel WF, Duszynski TJ, Yiannoutsos C, Halverson PK, et al. Symptoms and symptom clusters associated with SARS-CoV-2 infection in community-based populations: Results from a statewide epidemiological study. PLoS One 2021;16(3):e0241875 [FREE Full text] [doi: 10.1371/journal.pone.0241875] [Medline: 33760821]
27. Data API. The COVID Tracking Project. URL: https://covidtracking.com/data/api [accessed 2021-04-29]
28. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus pandemic (COVID-19). Our World in Data. 2020. URL: https://ourworldindata.org/coronavirus [accessed 2021-04-29]
29. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
30. Dixon BE, Caine VA, Halverson PK. Deficient response to COVID-19 makes the case for evolving the public health system. Am J Prev Med 2020 Dec;59(6):887-891 [FREE Full text] [doi: 10.1016/j.amepre.2020.07.024] [Medline: 32978011]

Abbreviations

ETL: extract, transform, load
IU: Indiana University
PT: Pacific Time
REDCap: Research Electronic Data Capture
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Eysenbach; submitted 05.10.20; peer-reviewed by H Luo, A Chen; comments to author 10.11.20; revised version received 14.12.20; accepted 14.04.21; published 20.05.21.

Please cite as:
Dixon BE, Mukherjee S, Wiensch A, Gray ML, Ferres JML, Grannis SJ
Capturing COVID-19–Like Symptoms at Scale Using Banner Ads on an Online News Platform: Pilot Survey Study
J Med Internet Res 2021;23(5):e24742
URL: https://www.jmir.org/2021/5/e24742
doi:10.2196/24742
PMID:33872190

©Brian E Dixon, Sumit Mukherjee, Ashley Wiensch, Mary L Gray, Juan M Lavista Ferres, Shaun J Grannis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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 App-Based Intervention to Support First Responders and Essential Workers During the COVID-19 Pandemic: Needs Assessment and Mixed Methods Implementation Study

Stacie Vilendrer[1], MD, MBA, MS; Alexis Amano[1], MS; Cati G Brown Johnson[1], PhD; Marissa Favet[2]; Nadia Safaeinili[3], MPH; Jacqueline Villasenor[4]; Jonathan G Shaw[1], MD, MS; Attila J Hertelendy[5,6], PhD; Steven M Asch[1,7], MD, MPH; Megan Mahoney[1], MD

[1]Division of Primary Care and Population Health, Stanford University School of Medicine, Stanford, CA, United States

[2]College of Liberal Arts and Sciences, University of Iowa, Iowa City, IA, United States

[3]Health Policy and Management, University of California, Berkeley, Berkeley, CA, United States

[4]John Muir College, University of California, San Diego, La Jolla, CA, United States

[5]Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

[6]Disaster Medicine Fellowship, Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[7]Center for Innovation to Implementation, Veterans Affairs, Menlo Park, CA, United States

Corresponding Author:
Stacie Vilendrer, MD, MBA, MS
Division of Primary Care and Population Health
Stanford University School of Medicine
1265 Welch Road
Mail Code 5475
Stanford, CA, 94305
United States
Phone: 1 650 736 5211
Email: staciev@stanford.edu

Abstract

Background: The COVID-19 pandemic has created unprecedented challenges for first responders (eg, police, fire, and emergency medical services) and nonmedical essential workers (eg, workers in food, transportation, and other industries). Health systems may be uniquely suited to support these workers given their medical expertise, and mobile apps can reach local communities despite social distancing requirements. Formal evaluation of real-world mobile app–based interventions is lacking.

Objective: We aimed to evaluate the adoption, acceptability, and appropriateness of an academic medical center–sponsored app-based intervention (COVID-19 Guide App) designed to support access of first responders and essential workers to COVID-19 information and testing services. We also sought to better understand the COVID-19–related needs of these workers early in the pandemic.

Methods: To understand overall community adoption, views and download data of the COVID-19 Guide App were described. To understand the adoption, appropriateness, and acceptability of the app and the unmet needs of workers, semistructured qualitative interviews were conducted by telephone, by video, and in person with first responders and essential workers in the San Francisco Bay Area who were recruited through purposive, convenience, and snowball sampling. Interview transcripts and field notes were qualitatively analyzed and presented using an implementation outcomes framework.

Results: From its launch in April 2020 to September 2020, the app received 8262 views from unique devices and 6640 downloads (80.4% conversion rate, 0.61% adoption rate across the Bay Area). App acceptability was mixed among the 17 first responders interviewed and high among the 10 essential workers interviewed. Select themes included the need for personalized and accurate information, access to testing, and securing personal safety. First responders faced additional challenges related to interprofessional coordination and a “culture of heroism” that could both protect against and exacerbate health vulnerability.

Conclusions: First responders and essential workers both reported challenges related to obtaining accurate information, testing services, and other resources. A mobile app intervention has the potential to combat these challenges through the provision of disease-specific information and access to testing services but may be most effective if delivered as part of a larger ecosystem of

XSL•FO
RenderX

support. Differentiated interventions that acknowledge and address the divergent needs between first responders and non–first responder essential workers may optimize acceptance and adoption.

(J Med Internet Res 2021;23(5):e26573) doi:10.2196/26573

KEYWORDS

COVID-19; pandemic; health literacy; social media; quality improvement; police; emergency responders; physicians; disasters; natural disasters; health behavior; literacy; app; intervention; adoption; accessibility; usability; support; testing

Introduction

The COVID-19 pandemic has presented unique obstacles to first responders as well as to other essential workers who ensure that health and other basic needs of the public continue to be met [1,2]. Health systems may be uniquely positioned to support these individuals given their expertise and position embedded within local communities. Efforts to understand the challenges of first responders and essential workers and to evaluate support of these groups by health systems during a pandemic response are needed.

The definition of a "first responder" varies between state and federal agencies but commonly encompasses emergency medical personnel, firefighters, and law enforcement officers [3]. Their duties can lead to adverse mental, physical, and social consequences [4]. Although academic attention has focused on the intense mental and emotional strain on health workers, such as during the severe acute respiratory syndrome (SARS) epidemic [5,6] and the COVID-19 pandemic [7-9], nonmedical first responders also face adverse mental health consequences [10,11]. Other adverse effects include worker absenteeism [12,13], an exacerbation of predisaster socioeconomic inequalities [14], and an increased risk of contracting the disease itself in an outbreak setting [15,16].

Less is known about the impact of disasters on essential workers not otherwise considered to be first responders, defined by their work in food and agriculture, construction, transportation, and other sectors of the economy deemed essential [17,18]. COVID-19 has spurred research in this population, with documented increased rates of depression and anxiety [19] and substance abuse [20] during the pandemic. Most concerning, the relative economic vulnerability of essential workers, including their need to leave home to work, further exacerbates existing socioeconomic and racial disparities seen in COVID-19 morbidity and mortality [17,21-23].

Certain factors can mitigate the stress experienced by first responders and essential workers [20]. Knowledge about a disease process may be protective, as essential workers without a health care background appear to experience a higher level of distress than workers who have training in disease processes [24]. Open, reliable information channels—even in a setting with many unknowns—may also be protective [25]. To date, documented efforts to support these workers have focused on prioritized access to testing [26], mental health support [18], and other health resources [27,28], but not necessarily on access to information.

Dissemination of this information occurs through formal and informal networks during a pandemic response [29]. However, the challenges brought on by the COVID-19 pandemic have highlighted a pre-existing disconnect between the health system and public health infrastructure [30,31]. Without these partnerships, first responders and essential workers cannot benefit from the expertise of the health systems located right in their communities. Given the uncertainties inherent in an unfolding pandemic, additional channels that draw upon health system expertise may be needed.

To this end, digital technologies, including mobile apps, appear to provide unique support in the context of social distancing efforts during a pandemic [32-34]. Formal evaluations of digital efforts to support first responders and essential workers are lacking; however, optimizing these upstream interventions may help mitigate health disparities and other adverse effects associated with their roles during a pandemic.

Methods

We conducted a qualitative user needs assessment and implementation evaluation (outcomes including adoption, appropriateness, and acceptability) [35] of an academic medical center–sponsored app-based intervention designed to support the information and testing needs of first responders and essential workers early in the COVID-19 pandemic. We further used app download data to triangulate [36] our understanding of the adoption of the app among the target populations using a convergent [37] mixed methods approach.

Setting

In April 2020, a quaternary academic medical center, Stanford Health Care (SHC, Palo Alto, California, United States) in partnership with a large technology company (Apple, Cupertino, California, United States) launched the *Stanford Medicine COVID-19 Guide for First Responders and Essential Workers* app (COVID-19 Guide App) with the goal of providing evidence-based information related to COVID-19, screening guidance, and access to testing services to support 1.1 million first responders and essential workers in the San Francisco Bay Area [38-40]. The dissemination effort by the institution included emails to local first responder and essential worker representatives, facilitated in part through previously established partnerships between SHC emergency physicians and emergency medical services (EMS) and fire organizations within the community developed through the institution's EMS Fellowship. This study was conducted by an interdisciplinary evaluation group led by the Stanford Medicine Evaluation Sciences Unit. First responder interview participants were recruited from Bay Area county police, fire, and EMS departments and among physicians connected to the health system's EMS physician fellowship. Essential worker interview participants were

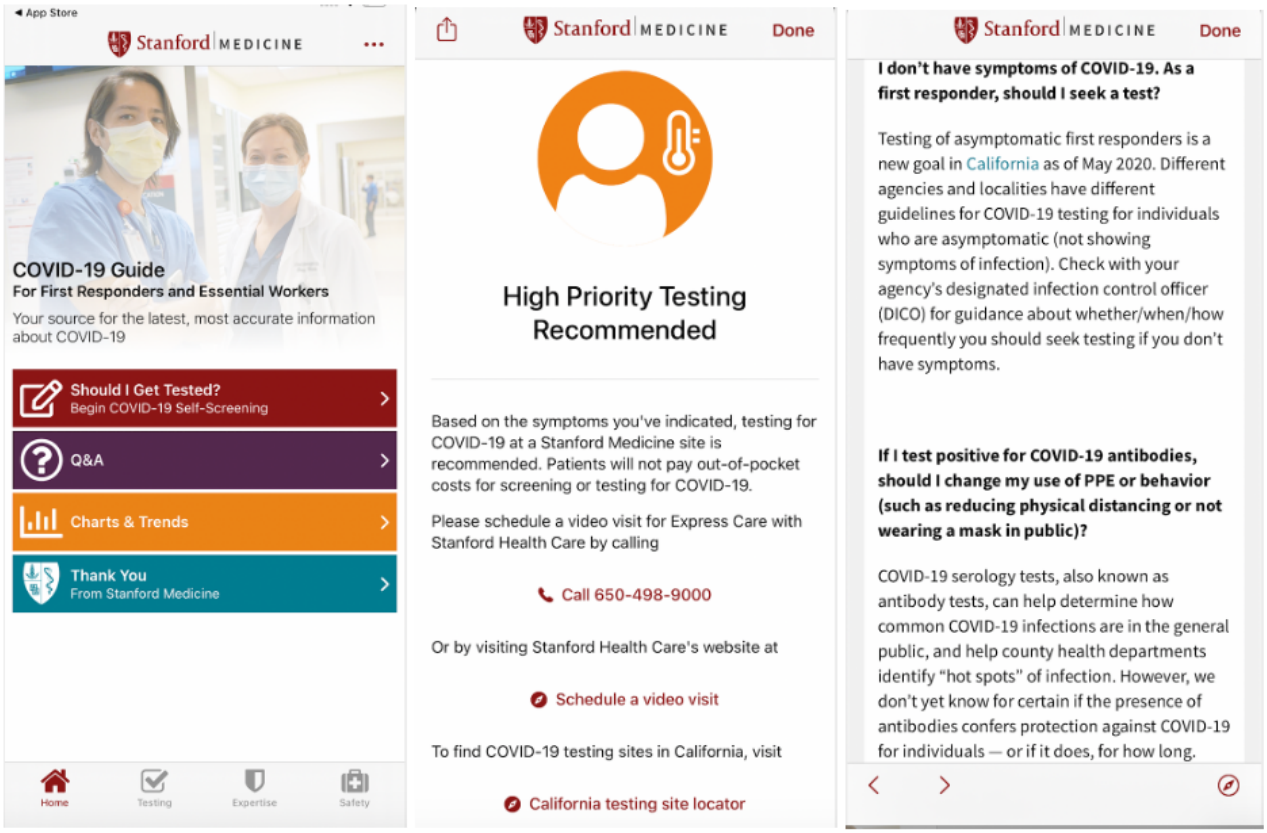
XSL•FO
RenderX

recruited from local grocery stores, gas stations, care facilities, and civil infrastructure support businesses.

Intervention

The COVID-19 Guide App includes a screening questionnaire that provides testing recommendations with links to access further testing and care, COVID-19 resources from Stanford Health Care physicians and scientists, and foundational information on viruses and prevention (Figure 1) [38].

Figure 1. Example user displays of the Stanford Medicine COVID-19 Guide for First Responders and Essential Workers app.



Sample

First responder participants were drawn from police, fire, and EMS departments in Alameda, San Mateo, San Francisco, and Santa Clara counties through email and telephone call outreach to local departments. All fire personnel are cross-trained in EMS, although not all EMS personnel have fire training [41]; participants from both groups were recruited. Emergency medicine physicians who partnered with local fire and police departments through the EMS Fellowship [42] at Stanford School of Medicine were also interviewed using a convenience sampling approach. Snowball sampling was leveraged across all categories to connect with additional participants in harder-to-reach groups. Interviews were conducted until thematic saturation within each professional subgroup was reached.

Essential workers were recruited using a convenience sampling approach in which researchers sought interviews in person at local businesses. When a paucity of Spanish-speaking essential workers was recognized, we shifted to a telephone-based convenience sampling approach targeted at this population to better reflect demographics of the region [43] until thematic saturation for both English- and Spanish-speaking essential workers was reached. Pre-existing experience with the app was not a requirement for participation, and the interview protocol itself accommodated both prior experience and no prior experience with the app (Multimedia Appendix 1).

Qualitative Data Collection

From May to September 2020, 30-minute semistructured interviews using a protocol grounded in the Consolidated Framework for Implementation Research were conducted by multiple authors (SV, NS, AA, MF, CBJ) (Multimedia Appendix 1) [44]. Interviews were conducted with two or more researchers present, with rare exceptions for note-taking, training, and analytic purposes. First responder interviews were conducted by telephone and videoconference; essential worker interviews were conducted by telephone and in person. For Spanish speakers, a professional interpreter was enlisted to interpret interviewer questions and interviewee responses. First responder interviews were tracked confidentially, recorded, and transcribed verbatim using Rev [45]. Essential worker interviews were not recorded to protect the anonymity of these more vulnerable interviewees in their work-based settings. For these interviews, near-verbatim field notes were taken, and two or more researchers who were present at each interview contributed to the field notes in a debrief discussion immediately following the interview.

XSL•FO
RenderX

Quantitative Data Collection

Descriptive quantitative data, including the total number of downloads at the conclusion of the evaluation on September 1, 2020, were drawn from the app administrator account. User characteristics, including demographic and use data, were not available due to restrictions in data use agreements.

Qualitative Analysis

Interview transcripts were analyzed using an inductive rapid analytic process in which two researchers coded all interviews (SV, AA) [46,47]. We used an inductive-deductive approach to identify a priori themes related to access to information and testing services and emergent themes. Excerpts were presented in a matrix format following rapid analytic procedure methodology [48]. The matrix was populated with notes and quotations (AA) and reviewed by qualitative team members (SV, AA, MF) to reach consensus and compare themes across worker subgroups [49]. The Implementation Outcomes Framework guided the evaluation and presentation of implementation outcomes (appropriateness, acceptability, and adoption) [50].

Throughout the assessment, early findings were reported back to operational leads to inform ongoing improvement using the Lightning Report Method described in 2019 by Brown-Johnson et al [51]. This project was reviewed by the Stanford Institutional Review Board and did not qualify as human subjects research (Protocol ID 56126). All participants gave verbal consent to be interviewed, and first responders gave consent to be audio-recorded.

Results

Participant Characteristics

A total of 27 interviews were conducted with 17 first responders (63%) and 10 essential workers (37%) (Table 1). The number of participants within each subcategory was comparable, including 5 police (29%), 7 emergency medical technician (EMT) (41%), 4 EMT with fire training (EMT/Fire, 59%), and 5 physician police/EMS advisor (PoliceMD/FireMD, 29%) interviewees in the group of 17 first responders and 5 food and agriculture (50%) and 5 transportation and civil infrastructure (50%) interviewees in the group of 10 essential workers. The food and agriculture subgroup included 4 grocery workers and 1 dining supervisor, while the transportation and civil infrastructure subgroup included 3 gas attendants, 1 bus driver, and 1 electrician. First responders tended to be male (16/17, 88%), 30-49 years old (11/17, 65%), and White (10/17, 59%), and most reported working in their field for more than 10 years (10/17, 59%). The essential workers were 50% male (5/10), of varying ages, mostly Hispanic/Latino (7/10, 70%), and had worked in the field for either less than 5 years (4/10, 40%) or more than 10 years (5/10, 50). The primary language of half of the essential workers (5/10, 50%) was Spanish, whereas all first responder interviews were conducted in English.

XSL•FO
RenderX

Table 1. Participant demographics of the first responders and essential workers (N=27).

Variable	Value, n (%)	
	First responders (n=17)	Essential workers (n=10)
Role		
Police	5 (29)	N/A[a]
EMS[b] overall	7 (41)	N/A
EMS with fire training	4 (59)	N/A
Physician police/EMS advisor	5 (29)	N/A
Food and agriculture	N/A	5 (50)
Transportation and civil infrastructure	N/A	5 (50)
Gender		
Male	15 (88)	5 (50)
Female	2 (12)	5 (50)
Age (years)		
20-29	2 (12)	1 (10)
30-39	7 (41)	2 (20)
40-49	4 (24)	2 (20)
50-59	3 (18)	1 (10)
60-69	1 (6)	3 (30)
70-79	0 (0)	1 (10)
Years working in occupation		
0-4	1 (6)	4 (40)
5-9	6 (35)	1 (10)
10+	10 (59)	5 (50)
Race/ethnicity		
Hispanic/Latino	5 (29)	7 (70)
American Indian or Alaskan Native	0 (0)	0 (0)
Asian or Asian American	2 (12)	1 (10)
Black or African American	0 (0)	0 (0)
Native Hawaiian or other Pacific Islander	0 (0)	0 (0)
White	10 (59)	0 (0)
Other/multiracial	0 (0)	2 (20)
Primary language		
English	17 (100)	5 (50)
Spanish	N/A	5 (50)

[a]N/A: not applicable.

[b]EMS: emergency medical services.

Implementation Outcomes of the COVID-19 Guide App

Implementation outcomes of the COVID-19 Guide App differed between the first responder and essential worker groups. Perspectives on the app varied by group, with moderate to low adoption, mixed appropriateness, and mixed acceptability seen in first responders, and low adoption, high appropriateness, and high acceptability seen in essential workers.

Adoption

The COVID-19 Guide App received 8262 views from unique devices, from which 6640 individuals downloaded the app (80.4% conversion rate) from the launch date of April 7, 2020 to the end of the evaluation period on September 1, 2020. Assuming all downloads were in the target demographic within

XSL•FO
RenderX

the Bay Area, this suggests a 0.61% adoption rate of the region's 1,093,781 first responders and essential workers [39,40].

In the qualitative interviews, 2 participants had used the app and were able to share their experience directly. The majority were not familiar with the app but shared their perspectives based on a short description of the app.

Moderate to low adoption was described in the first responder group. First responder interviews shed light on the underlying reasons limiting widespread adoption. Physicians largely felt their involvement in the app development occurred "fairly late in the game" (FireMD 1, 2), perhaps due to the poor visibility of the emergency fellowship program within the institution (FireMD 1, 4).

In the essential worker group, on the other hand, none of the interviewees reported direct experience with the app, although their attitudes toward a theoretical app that provided COVID-19 testing were positive (see the *Acceptability* section below).

Appropriateness

First responders reported mixed perceptions of the appropriateness of the COVID-19 Guide App. All interviewees reported owning smartphones, with most using mobile apps for work. Despite the seeming appropriateness of an app-based solution for this group, some reported that the features of this app were less relevant for first responders. Although some thought the screening features and link to testing resources were personally helpful (PoliceMD 3, EMT/Fire 2) or useful for patients (EMT/Fire 7), others felt these features were already covered by other processes in their organization. Some reported frustration with the app's inability to link a positive symptom screen to a scheduled test due to technical and privacy limitations; instead, the app would "take [them] to the end of a little pathway that would be like, 'You should call your [department infection control officer] to get tested.' No [kidding]." (FireMD 2).

Conversely, essential workers reported high appropriateness of the COVID-19 Guide App. Similar to first responders, all essential workers owned smart phones, and an overwhelming majority used mobile applications for work, indicating an ability to access and navigate mobile applications.

Acceptability

First responders expressed mixed acceptability of the app. A few participants who had directly used the app had positive perceptions (EMT/Fire 3, Police 1), sharing benefits of "quick answers" that were "super easy to access" (EMT/Fire 3).

Additionally, first responders also noted the need for a succinct, trusted data source (EMT/Fire 2), such as this app, particularly one that "agreed" with county information (Police 3, FireMD 4, PoliceMD 1). Indeed, brevity and trusted branding was seen as highly valuable and missing after a midimplementation shift to incorporate an outside partnership to produce content: "…Now it's just too hard to filter what is coming as sort of medical recommendation versus what almost seemed like an ad" (FireMD 4).

Although no essential workers reported direct experience with the app, their attitudes toward a theoretical app that provided COVID-19 testing and disease information were overwhelmingly positive: "I'm looking forward to this app" (Grocery 1).

Like first responders, essential workers reported a desire for succinct, up-to-date, and trustworthy information (Grocery 4, Food Services 5). Essential workers noted that ideally, information would be presented in multiple languages to address reported widespread misinformation, particularly in non–English-speaking communities (Gas 2, Electrician 4). This information would also ideally be delivered through social media channels by physicians or nurses, as trust in the health system was reportedly higher than in media or government.

Challenges Faced by First Responders and Essential Workers

First responders and essential workers experienced a number of related challenges, described in the following qualitative themes: a need for personalized information with no clear playbook, parsing information takes time, misinformed beliefs fuel conflict, priority testing gets them back to work, fear of bringing an amorphous enemy home, and a lack of resources to fight the pandemic. First responders faced additional challenges related to the need for interprofessional coordination and a double-edged "culture of heroism" that could both protect against and exacerbate mental and health vulnerabilities (Table 2).

XSL•FO
RenderX

Table 2. Impact of the COVID-19 pandemic on first responders and essential workers.

Challenges	Illustrative quotations	
	First responders	Non–first responder essential workers
Need for personalized answers with no clear playbook	• "When COVID hit, I was taken out of my primary role… and put into a COVID response detail" (Police 1) • "And early on it was a lot of questions [from EMS[a]] 'this patient … didn't tell us she had a fever but when they got to the hospital, she had a fever and she turned out to have been on the Princess Cruise' [early COVID-19 outbreak] …it was tough because there wasn't a clear playbook for so many different situations." (Fire MD 4)	• "Before, food was served in a dining room…now residents eat in their own rooms. We set up dining carts to bring food to residents, which is exhausting." [paraphrased] (Food Services 5) • "I want information in Spanish" (Gas 2) • "I think it would be helpful to have more knowledge on herbal treatments for COVID… I prefer to use this" (Electrician 4)
Parsing information takes time	• "I pretty much rely on my division chief to disseminate information, because he spends all day reading and looking at case studies and such." (EMT[b]/Fire 7)	• "We are being bombed by so much information everywhere, so it is hard to know what is true." (Bus Driver 5)
Misinformed beliefs fuel conflict	• "The amount of misinformation has been pretty surprising. So, I don't know that our fire departments are affected more heavily than anyone else, but there's a lot of suspicion and a lot of misinformation." (FireMD 2) • "We've had a couple people who have come up and said, 'COVID-19 is a conspiracy. It's not real.'…. just acting as though we're propagating this lie, which, as firefighters, we're not used to [such] conflict with our community members…It's a little bit disappointing to see people, A: deny it and B: be confrontational with us." (EMT/Fire 3)	• "We have to deal with [people not complying with mask policies] every single day… A lot of people have the mask on but around their neck" (Bus Driver 1) • "I don't trust information from the government. I don't think it is accurate or reliable. The number of sick people is overexaggerated… they are trying to scare people." (Food Services 5) • "A German doctor is leading the work around chlorine dioxide to balance pH which improves immunity…there are some websites that are against chlorine dioxide but many that show that it works, so it is a matter of perspective what you believe" [paraphrased] (Electrician 4)
Priority testing gets them back to work	• "But yeah, there hasn't been much else as far as support, like where we could get tested, information on how we get tested." (EMT 1) • "I didn't want anybody [in law enforcement] to get sick at work from it, and I didn't want community members to get sick from my teams. So I would just err and put them off from work until we heard back, but the priority testing… is really the only option I have [to get them back to the field]" (Police 2)	• "[My employer] didn't give us a place to get tested… I wish that we had more access to testing for drivers… We drive every day, for 5 days a week and we take care of all kinds of people" (Bus Driver 5)
Fear of bringing an amorphous enemy home	• "[Police officers] have a lot of anxiety about bringing it home to their families and that sort of thing." (PoliceMD 3) • "The thing that obviously is concerning to people is the last thing we want to do is bring this thing home, right? Especially those of us who have small kids …and spouses." (Police 4)	• "It's been very tough. I need to work even though I don't feel comfortable doing so" (Grocery 4) • "At first, I couldn't breathe. The first two weeks, I was panicked. What happens if I die? I stopped reading the news, and it's getting much better now." (Gas 1)
Lack of resources to fight the pandemic	• "I think what we're having that's been a disappointment is the lack of PPEs[c]. And then the people that have been price gouging with the PPEs… And then we get them, and they're not N95s. So… we can't use them in the field." (EMT 6) • "…I mean our city fire departments are getting… [mixed] messages of 'you're heroes, but the city's not going to make nearly as much money this year as they normally do, so we're going to slash your budget.'" (FireMD 2)	• "We have coworkers that already died, we have coworkers they already have COVID. Now they give us a set of masks, a little thin piece of material. Come on." (Bus Driver 5)

XSL•FO
RenderX

Challenges	Illustrative quotations	
	First responders	Non–first responder essential workers
Need for interprofessional coordination (first responders only)	• "And I can say that we've [transported] 13 positive COVID cases. I've only been notified on two of them." (EMT/Fire 5) • "Where to park, where to go, what the entrance point is, what's the exit point [at the hospital], is there a priority line? If law enforcement comes in, and they need to do something, do they have to wait in the general line and wait behind 10 people, or are they going to be let in a different way or checked in quickly, or given priority?" (Police 4)	Did not emerge in this group
Double-edged culture of heroism (first responders only)	• "There's a certain expectation that we hold ourselves to when we come to work and put a uniform on. Regardless of the environment or whatever's presented…we just do our best to protect ourselves. Are some of us fearful of it [COVID-19]? Sure, but it doesn't create necessarily anxiety." (Police 5) • I think that because [police officers] tend to be the typical first responders, they hold things in … so they don't deal with anxiety about these sorts of issues…They only reach a crisis point before they admit it, is what tends to happen." (PoliceMD 3) • "…I think that it was tricky because I felt like we were testing pretty much anybody we wanted to at Stanford if we had a suspicion, but it was harder to tell a firefighter, 'Hey, you probably should get tested. You had a high risk exposure.'" (FireMD 1)	Did not emerge in this group

[a]EMS: emergency medical services.

[b]EMT: emergency medical technician.

[c]PPE: personal protective equipment.

Need for Personalized Answers With No Clear Playbook

First responders and essential workers reported that throughout the pandemic, they found themselves adjusting to new protocols and professional responsibilities, which came with many uncertainties.

First responders wanted to know how standard work protocols should change in the setting of the COVID-19 pandemic, particularly in the setting of limited personal protective equipment (PPE) resources (FireMD 1). These protocol questions impacted standard procedures, such as transporting evidence from an out-of-state crime scene involving suspects with confirmed COVID-19 (PoliceMD 5), as well as human resource and personal health concerns, such as contemplating how to manage COVID-19 risk for pregnant staff and officers (Police 3). In the absence of any precedent, professionals were forced to make challenging decisions with limited information.

Essential workers also wanted personalized information (Gas 2, Electrician 4, Gas 1), though these needs were not typically related to altered professional responsibilities but were instead related to language, personal preferences, and need for increased financial support. Both groups wanted more local dynamic information about COVID-19–positive cases (Gas 3, Police 4, EMT/Fire 5).

Parsing Information Takes Time

Members of both groups acknowledged an overabundance of COVID-19 information through the media, internet, social media, and their professional settings. This aspect was particularly challenging given the dynamic nature of the available evidence (EMT/Fire 3, Police 2). Although the first responders appreciated county-level dashboards, professional daily emails, and Incident Command Center communications that made this information easier to digest (Fire/MD 4, EMT/Fire 5, Police 3, Police 5), others acknowledged significant time spent by either themselves (Gas 2, Grocery 3, Police 1) or someone in their work setting to digest the information (EMT/Fire 7).

Although first responders received their information largely from professional networks, essential worker information sources leaned toward social media (Electrician 4) and other websites (Grocery 2). The quantity of data led to skepticism:

> *Is there any [accurate data]? I feel so in the blind. I feel like there's information flying from every direction and so much of it is unreliable.* [EMT/Fire 7]

Misinformed Beliefs Fuel Conflict

The lack of reliable data sources seemed to fuel mistrust and misinformation. Both groups expressed an explicit distrust of media (EMT 4, Bus Driver 5) and government (Food Services 5). Workers also reported encountering community members

XSL•FO
RenderX

through their work who were misinformed, many of whom were reported to believe that the pandemic was a conspiracy (EMT/Fire 3). Lapses in community cooperation with mask requirements created novel conflict on the job, including the need to address "hostile" behavior in the community (Gas 1, EMT/Fire 3).

Priority Testing Gets Them Safely Back to Work

Both groups acknowledged some challenges in scheduling testing both through the app and in general through a work referral or on their own. Once a test was scheduled, however, experiences were positive. Interviewees emphasized that prioritizing testing was critical for enabling first responders and essential workers to return to their duties. In first responder organizations in which testing access was straightforward, this was thought to be due to designated COVID-19 support within the organization (Police 5, Electrician 4, EMT/Fire 5).

By contrast, essential workers reported frustration with accessing testing (Bus Driver 5). In the majority of cases where testing support was not available through work, participants sought alternate options through their primary care providers or the county (Bus Driver 5, Grocery 4, Gas 2); however, 1 participant was unsure of how to access testing at all (Gas 3).

Fear of Bringing an Amorphous Enemy Home

The stressful effects of the pandemic were salient across several interviews in both groups. These concerns centered on the amorphous nature of the disease (EMT/Fire 2), a need to go to work despite not feeling safe (Grocery 4), and fear of bringing the virus home to family (PoliceMD 3, EMT 4, Grocery 4, Food Services 5). Early high mortality rates documented in other countries contributed to widespread "paranoia" (EMT/Fire 5, Electrician 4), and reading the news worsened anxiety for some (Gas 1, Grocery 1).

The nature of a pandemic also meant that the workers' families were affected, demanding their attention on top of their existing professional duties (FireMD 4). The prolonged and unpredictable nature of the pandemic was described as a "roller coaster" (EMT/Fire 7) with a "hurry up and wait" pace (EMT/Fire 2). One responder had "never been so scared and so bored in my life" (EMT/Fire 2). The mental suffering that resulted from the COVID-19 pandemic impacted both first responder and essential worker communities, although open discussions related to mental health were described as "muted" (Bus Driver 5). Although no first responders acknowledged wanting additional mental health support (see "Culture of Heroism" below), at least one essential worker acknowledged she did not have the support she needed (Food Services 5).

Lack of Resources to Fight the Pandemic

A lack of resources emerged as a dominant theme in conversations for both first responders and essential workers. Several participants noted insufficient quantities of PPE, to the point that police officers were reported to "go begging" at one point to obtain a sufficient supply (PoliceMD 3). Public budget cuts and subsequent staffing shortages were also cited as major challenges (Police 4). Although some workers belonged to unions, which made some efforts to support workers, their impact was perceived to be limited (Bus Driver 5, Grocery 4).

The broader economic impact of the pandemic also shifted the workers' financial situations. Although some felt that the essential nature of their jobs made them more secure (Grocery 1, 2), this sentiment did not extend to everyone, as essential workers reported being furloughed (Bus Driver 5) or felt insecure in their position (Gas 1, 2). Finding financial support in the case of illness was a major challenge (EMT 1), particularly for workers who were new or did not have sick days (EMT/Fire 7). For this reason, some cited a reluctance to get tested because they were unable to take two weeks off work (EMT 6). Essential workers, more than first responders, cited financial insecurity as a major ongoing challenge (Grocery 4, Food Services 5).

Need for Interprofessional Coordination (First Responders Only)

Some themes emerged that were specific to first responders, including the need for interprofessional coordination. Health system support of first responders was felt to be critical, given their role as "the tip of the spear" (EMT/Fire 5). Specific areas for coordination with health systems included follow-up information regarding a transported patient's health status. When these communication channels fail, "anxiety increases significantly," as first responders wonder about potential exposure (EMT/Fire 5).

Another challenge emerged relating to hospital entry to respond to a crisis while bypassing the COVID-19 fever checkpoint (Police 4). In this case, the participant pointed out that because the other professional organizations already screen their workers before they start their shifts, the health system should be able to skip this step. Finally, coordination of COVID-19 information across health systems and counties up the Incident Command System was also cited as a challenge (Police 3, FireMD 4, PoliceMD 5).

Double-Edged Culture of Heroism (First Responders Only)

Several first responders shared their ongoing willingness to face the pandemic despite unfavorable circumstances because they had made a commitment to serve when choosing their profession (EMT 1, 4; EMT/Fire 3, 14; Police 1, 4, 5). Some described an underlying desire to be helpful (EMT/Fire 2) and an expansive view of their own roles in the community (Police 4). The role of choice in their profession was reported to protect against the adverse mental health effects of their roles, as with the attitude that "this is what I signed up for" (EMT/Fire 3).

However, this "cavalier" culture also led to a reluctance to report symptoms and obtain health services (PoliceMD 3). Even physician advisors shared reluctance to tell first responders they needed testing (FireMD 1). One EMT/Fire responder noted that few first responders recognized that their own colleagues may be struggling and undergoing therapy to handle their mental stress (EMT/Fire 2).

XSL•FO
RenderX

Solutions to Challenges in the Field

Participants provided several solutions to address these challenges, some of which they thought could be delivered through an app (Table 3). To address misinformation and fear of SARS-CoV-2, several participants validated the potential benefit of succinct, accurate information delivered through an app. Others thought that sharing accurate information through social media would be most effective given the increased visibility of these platforms among essential workers. Yet others felt the need for "higher touch" strategies, possibly including holding local question and answer (Q&A) sessions (FireMD 2). The personalization of information in any solution was a recurrent theme:

> *So to me that [personalized information] seems more in line with our overall philosophy at Stanford of precision medicine. It's more like precision information because then it's filtered, it's more specific to you. You don't get confused when you're reading about stuff that doesn't apply to you.* [FireMD 4]

In the absence of precision information, many solutions were developed in the field: "Some of the crews just kind of came up with stuff on their own" (FireMD 4). Other essential workers found support in managing conflict with the public from their employers (Grocery 4) or through their own volition. One bus driver reported company instructions to call a hotline when riders were not compliant with wearing masks, but she instead refused to drive the bus: "I use my [face covering gesture] until they put it on" (Bus Driver 5).

Expanding the physician-first responder partnerships to a broader audience, particularly to the police, where only informal relationships existed, was suggested. A specific request was for a designated number or person within the health system to call when the COVID-19 status of a transported patient was under question.

Personalized information also encompassed health system–specific COVID-19–related information. Some first responders expressed a desire for health systems to be forthcoming about their own infection control practices and current rates: "That's where you just put the notion of, 'Hey, we're [health system] safe…You can transport patients to our hospitals and we won't get you contracted'" (EMT 4). First responders also wanted detailed, localized COVID-19–positivity data, expressing a desire to better understand the geographic spread of cases throughout the county (Police 4, EMT/Fire 2, 7).

XSL•FO
RenderX

Table 3. Participant-derived digital and nondigital solutions to support first responders and essential workers during a pandemic.

Challenge	Participant-derived digital and nondigital solutions	Example quotations
Need for personalized answers with no clear playbook	• Physician–first responder department partnerships (Fire MD 1, 4, Police MD 3, 5, Police 1, 3, 5, EMT[a]/Fire 7) • Physician-led Q&A[b] sessions (Fire MD 2) • Colocation of county officer (Police 2)	• "[E]specially in our early stages of our Incident Command Center, we created an entire team called a Virus Response Team… at the end of the day… they're just your everyday cop… I think it would have been helpful to have guidance from somebody a little bit more knowledgeable in the field versus just picking up the information from the distributor and figuring it out from there." (Police 3)
Parsing information takes time/misinformed beliefs fuel conflict	• Physicians coming into work huddles to provide relevant, accurate COVID-19 information (Grocery 2, 4, Police 5) • Provide information on an app (Grocery 1, 4, Gas 1, EMT/Fire 2, 3, EMT 4, Police 4, Electrician 4, Food Services 5) • Provide information and/or advertise app on social media (Electrician 4), telenovelas (Food Services 5)	• "Maybe coming and visiting. There are two huddles a day—maybe you [health workers] could pop in during those." (Grocery Worker 2) • "I think that's a large part of my role is that for the most part, the personnel in my departments know me and so if they ask me a question and I answer a question, or if I bring something up, they basically trust most of what I say. Or at least have enough kind of history with me that they at least need to think about it... So, I mean part of my role is to be kind of a sounding board for what's true and what's probably not true" (FireMD 2)
Priority testing gets them back to work	• Provide comprehensive information on testing resources in the county (Police 1, FireMD 4) • Provide priority testing for first responders and essential workers (Police 1, 4, 5; Police MD 3)	• "I think just a formalized priority testing system for law enforcement has a huge impact on the organization… these are the essential workers that if they're not at work or they are sick, it affects teams of people, and cars, and community contact." (Police 1)
Fear of bringing an amorphous enemy home	• Provide resources for first responders to help them manage stress and anxiety (Police MD 3) • Ask the public to share whether someone is sick in the presence of emergency personnel (EMT/Fire 5)	• "[M]aking sure you've got reliable sources of information and things like that [can be helpful]…because it can just help identify that yes, you're having anxiety about this. Here's some tools to help with it." (PoliceMD 3)
A lack of resources to fight the pandemic	• Share PPE[c] and other resources when possible (EMT 4, 6) • Promote consistent mask use to the public (EMT/Fire 7, 3)	• "I guess just shortages on masks…It's our company that was supposed to provide for us, but it's even hard for us to get supplies sometimes." (EMT 4) • "Wearing masks and not getting other people sick. That's kind of supporting us... I've been telling people that have been asking me the best thing they can do is to keep themselves healthy, to not over impact our system by, all the things that everyone's being asked to do by washing their hands and covering their face when they're in public. That's certainly a factor." (EMT/Fire 7)
Need for interprofessional coordination (first responders only)	• Keep first responders updated on the outcomes of patients under investigation (EMT 6, EMT/Fire 5, 7) • Keep first responders informed on the process of entering hospitals to facilitate efficient emergency response (Police 4)	• "Hospitals should have their infectious disease control departments, and they should be responsible for reaching out to people when they make positive COVID results." (EMT 6)
A double-edged culture of heroism (first responders only)	• Physician–first responder department partnerships (PoliceMD 3) • Provide resources for first responders to help them manage stress and anxiety (Police MD 3)	• "They're not that kind of group that usually shares those things, unless you have a preexisting relationship with them. That can be hard to get to." (PoliceMD 3)

[a]EMT: emergency medical technician.

[b]Q&A: question and answer.

[c]PPE: personal protective equipment.

XSL•FO
RenderX

Heath Worker Partnerships at the Heart of a Community Solution

A number of participants desired a stronger connection to a health worker during COVID-19. The physician partners themselves felt they were often the best resource to support first responders during the pandemic: "we are the best conduit [through] the already established trust that we've built up through the years" (FireMD 1). These participants felt their role encompassed the development and maintenance of long-term, trusting relationships with these first responder communities (FireMD 1, PoliceMD 5). First responders corroborated the special nature of the relationship in times of crisis, as one physician partner was described as "a really good resource. He's been coming by the station now too, to talk with us and to check in on COVID related topics, as well as other [topics]" (EMT/Fire 7).

In addition to their formal physician partnerships, participants also received information through informal networks into the health care community, such as a former employee who worked in the health system at the time of the interview (Grocery 3). Those with shift-based work welcomed the idea of a health worker presence at the beginning of their shift: "That may be an opportunity…to have someone present during those briefings to talk on any new developments or anything COVID related" (Police 5, Grocery 2). The human, in-person element of this relationship was emphasized.

Discussion

Principal Findings

Digital initiatives such as the Stanford COVID-19 Guide App have the potential to equip first responders and other essential workers with accurate information and access to testing resources, although adoption may depend on targeted publicity and early involvement of stakeholders with pre-existing relationships within these communities. While first responders are often included under the broader category of essential workers during outreach efforts, our findings suggest they face both overlapping challenges—related to the need for personalized and accurate information, access to testing, and ongoing personal safety—and divergent challenges, such as the need for interprofessional coordination and a double-edged culture of heroism, seen in the first responder sample alone. Optimal strategies for supporting each group should therefore differ based on their divergent needs, channels for outreach, appetite for digital and relationship-based approaches, and relationship to the health system (Table 4).

Physician partnerships emerged as a key resource for first responder departments, and these relationships are recognized as a foundational underpinning of a pandemic response [52-54]. Formalizing and nourishing these relationships holds promise in strengthening the overall ability of the health system to support its community during a pandemic. Interprofessional coordination was also cited as a major challenge, hampered by interoperability challenges between the electronic health records of EMS and the receiving health system [55]. Given these challenges, a trusted point of contact (eg, emergency department nurse line) between EMS and the health system may be explored as an alternate solution in a pandemic setting.

Our findings also suggest that first responders see increased risk as an integral part of assuming their role and may therefore adapt more readily to a new pandemic. Conversely, non–first responder essential workers decidedly *do not* see their roles as inherently risky outside a pandemic scenario. Without the training and cultural reinforcement around managing increased risk as a part of standard work, essential workers may suffer disproportionately from the onset of an infectious disease pandemic. The contributions of these workers during a pandemic should therefore not be diminished, and prioritized health services and hazard pay may be justified [56].

Similarly, misinformation emerged as a widespread challenge among the broader community. Although first responders have at least a basic degree of medical training and serve as a source of information to laypersons during a pandemic [57], essential workers often lack this formal training. Health literacy, defined as the ability to acquire, understand and enlist health information [58] becomes essential during a pandemic as first responders and essential workers come into regular, direct contact with members of the community. The reported reliance of essential workers on social media for information is concerning given its link to the spread of unvetted information [59-61]. Efforts to build health literacy, not only through digital interventions but also through channels with widespread use among vulnerable populations, including essential workers and non-English speakers, are urgently needed.

We note that the early learnings described here must be interpreted with the context that the COVID Guide App was rapidly deployed in just under three months after the first reported COVID-19 case in the United States [62]. Balancing the accuracy of an intervention in achieving its aim with the goal of rapid deployment will no doubt remain a challenge during this pandemic and other disasters; we recommend a standing advisory board with representation from target populations that can be mobilized in the short term to inform intervention development. In the long-term, community-based participatory research and human-centered design are parallel and overlapping methodologies that emphasize user expertise and creative problem-solving, with documented success in vulnerable populations [63-65]. These tools may inform improved evolutions of community interventions.

XSL•FO
RenderX

Table 4. Digital and nondigital recommendations to support first responders and essential workers during a pandemic.

Consideration	Recommendations	
	First responders	Essential workers
High need areas	• Advice on work adaptation • Streamlined accurate and timely information • Interprofessional coordination around follow up of patient diagnostic status and logistics • Priority access to rapid testing • Steady access to PPE[a]	• Acceptable resources to access accurate and timely information • Priority access to rapid testing • Steady access to PPE • Assistance navigating financial resources
Primary channels for outreach	• Formal physician partnerships with first responder departments • Employers (eg, through county infection control officers)	• Employers • Social media • Telenovelas and other culture-specific channels
Role of an app-based intervention	Small role • Provide nonpublic up-to-date information on working with health systems directly (eg, priority building entrances, rapid testing schedule, contact information to receive outcome updates on patients under investigation)	Large role • Source of credible health information that builds health literacy (eg, topics such as "protecting yourself and your family," "viral transmission," "symptom screening")
Role of physician/health worker outreach	Large role • Provide guidance on safe work adaptations • Facilitate provision of rapid testing and follow-up medical support • Assist in the dissemination of relevant health information	Small role • Provide guidance on safe work adaptations • Present health information to local businesses
Role of local health system	Large role • Facilitate prioritized access to testing and follow-up health services • Provide follow-up information to emergency medical service workers regarding patient diagnostic status • Develop and disseminate streamlined protocols for entering and exiting healthcare buildings during a pandemic	Large role • Facilitate prioritized access to testing and follow-up health services • Directly address misinformation related to pandemic

[a]PPE: personal protective equipment.

Limitations and Future Work

Limitations of this study include a lack of systematic sampling across professions in the non–first responder essential worker population, which resulted from challenges accessing this group in the field. We are unaware of similar work reaching this vulnerable population, which suggests this sample is still of value. Further, Spanish-speakers predominated our non–English-speaking sample, which reflects local demographics but does not account for the full variety of languages spoken in the community [43]. Furthermore, qualitative data capture differed between the first responders (verbatim transcriptions) and essential workers (near-verbatim field notes). This decision was made to protect essential worker anonymity given their potentially vulnerable position near their place of work. Efforts to mitigate data loss from an absence of recordings included a dedicated note-taker and immediate debrief discussions between two or more researchers to expand data capture. Finally, our learnings are limited to our local community, as the Bay Area may not be representative of populations and resources available elsewhere.

Future work will benefit from increased attention to the variety of first languages that exist within the community, including non-Spanish languages. The incorporation of a mental health component to this digital solution, reported as a need here and evaluated in other settings [66,67], may also be an opportunity for future work. Finally, evaluating digital support systems for EMS personnel as they face an increased number of deaths in the field as a result of the pandemic [68,69]—supporting decisions related to triage and/or end of life care—is another important area for future research.

Conclusion

First responders and essential workers face shared challenges related to obtaining accurate information and testing services, and securing their personal safety. Digital interventions such as mobile applications have the potential to combat these challenges through the provision disease-specific information and access to testing services. Such solutions are likely to be most effective if delivered as part of a larger ecosystem of support, and with early and direct input from those in these professions to understand how best to meet their specific needs. Given varying challenges between first responders and non-first

XSL•FO
RenderX

responder essential workers, our results indicate that differentiated interventions may leverage shared insights while also acknowledging differences in their occupational requirements and culture.

Acknowledgments

The authors would like to acknowledge Jared Roberts, Michael Halaas, and Catherine Krna for their support of the COVID-19 Guide App and programmatic evaluation described here.

Conflicts of Interest

None declared.

Multimedia Appendix 1
First responder and essential worker semistructured interview protocols.
[DOCX File , 13 KB - jmir_v23i5e26573_app1.docx]

References

1. Awais SB, Martins RS, Khan MS. Paramedics in pandemics: protecting the mental wellness of those behind enemy lines. Br J Psychiatry 2021 Feb 23;218(2):75-76. [doi: 10.1192/bjp.2020.193] [Medline: 33092658]
2. Djalante R, Shaw R, DeWit A. Building resilience against biological hazards and pandemics: COVID-19 and its implications for the Sendai Framework. Prog Disaster Sci 2020 Apr;6:100080. [doi: 10.1016/j.pdisas.2020.100080]
3. Bricker L, Petermann T, Hines M, Sands J. The Legal Definitions of a First Responder. Transportation Research Board. 2013. URL: https://www.nap.edu/catalog/22451/the-legal-definitions-of-first-responder [accessed 2021-05-06]
4. The Dialogue: A Quarterly Technical Assistance Journal on Disaster Behavioral Healthffects of Trauma on First Responders. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/sites/default/files/dtac/dialogue-vol14-is1_final_051718.pdf [accessed 2020-10-21]
5. Bai Y, Lin C, Lin C, Chen J, Chue C, Chou P. Survey of stress reactions among health care workers involved with the SARS outbreak. Psychiatr Serv 2004 Sep;55(9):1055-1057. [doi: 10.1176/appi.ps.55.9.1055] [Medline: 15345768]
6. Styra R, Hawryluck L, Robinson S, Kasapinovic S, Fones C, Gold WL. Impact on health care workers employed in high-risk areas during the Toronto SARS outbreak. J Psychosom Res 2008 Feb;64(2):177-183 [FREE Full text] [doi: 10.1016/j.jpsychores.2007.07.015] [Medline: 18222131]
7. Paterlini M. On the front lines of coronavirus: the Italian response to covid-19. BMJ 2020 Mar 16;368:m1065. [doi: 10.1136/bmj.m1065] [Medline: 32179517]
8. Du J, Dong L, Wang T, Yuan C, Fu R, Zhang L, et al. Psychological symptoms among frontline healthcare workers during COVID-19 outbreak in Wuhan. Gen Hosp Psychiatry 2020 Apr;67:144-145 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.03.011] [Medline: 32381270]
9. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
10. DePierro J, Lowe S, Katz C. Lessons learned from 9/11: mental health perspectives on the COVID-19 pandemic. Psychiatry Res 2020 Jun;288:113024 [FREE Full text] [doi: 10.1016/j.psychres.2020.113024] [Medline: 32315874]
11. Stogner J, Miller BL, McLean K. Police stress, mental health, and resiliency during the COVID-19 pandemic. Am J Crim Justice 2020 Jun 26;45(4):1-13 [FREE Full text] [doi: 10.1007/s12103-020-09548-y] [Medline: 32837167]
12. Gershon RRM, Magda LA, Qureshi KA, Riley HEM, Scanlon E, Carney MT, et al. Factors associated with the ability and willingness of essential workers to report to duty during a pandemic. J Occup Environ Med 2010 Oct;52(10):995-1003. [doi: 10.1097/JOM.0b013e3181f43872] [Medline: 20881624]
13. Garrett AL, Park YS, Redlener I. Mitigating absenteeism in hospital workers during a pandemic. Disaster Med Public Health Prep 2013 Apr 08;3(S2):S141-S147. [doi: 10.1097/dmp.0b013e3181c12959]
14. Weber L, Hilfinger Messias DK. Mississippi front-line recovery work after Hurricane Katrina: an analysis of the intersections of gender, race, and class in advocacy, power relations, and health. Soc Sci Med 2012 Jun;74(11):1833-1841. [doi: 10.1016/j.socscimed.2011.08.034] [Medline: 21978634]
15. Maurer LR, Perez NP, Witt EE, Ortega G. Protecting our own: equity for employees as hospitals battle COVID-19. Health Equity 2020 Sep 01;4(1):394-396 [FREE Full text] [doi: 10.1089/heq.2020.0024] [Medline: 32999949]
16. Chersich MF, Gray G, Fairlie L, Eichbaum Q, Mayhew S, Allwood B, et al. COVID-19 in Africa: care and protection for frontline healthcare workers. Global Health 2020 May 15;16(1):46 [FREE Full text] [doi: 10.1186/s12992-020-00574-3] [Medline: 32414379]
17. The Lancet. The plight of essential workers during the COVID-19 pandemic. Lancet 2020 May;395(10237):1587. [doi: 10.1016/s0140-6736(20)31200-9]

XSL•FO
RenderX

18. Benhamou K, Piedra A. CBT-informed interventions for essential workers during the COVID-19 pandemic. J Contemp Psychother 2020 Jul 18;50(4):1-9 [FREE Full text] [doi: 10.1007/s10879-020-09467-3] [Medline: 32836379]
19. De BR, Balanzá- MV, Mota J. Depression, anxiety and lifestyle among essential workers: a websurvey from Brazil and Spain during the COVID-19 pandemic. JMIR Preprints. Preprint posted online on July 24, 2020. [doi: 10.2196/preprints.22835]
20. McKay D, Asmundson GJ. Substance use and abuse associated with the behavioral immune system during COVID-19: the special case of healthcare workers and essential workers. Addict Behav 2020 Nov;110:106522 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106522] [Medline: 32622026]
21. McCormack G, Avery C, Spitzer AK, Chandra A. Economic vulnerability of households with essential workers. JAMA 2020 Jul 28;324(4):388-390 [FREE Full text] [doi: 10.1001/jama.2020.11366] [Medline: 32556217]
22. Rogers TN, Rogers CR, VanSant-Webb E, Gu LY, Yan B, Qeadan F. Racial disparities in COVID-19 mortality among essential workers in the United States. World Med Health Policy 2020 Aug 05;12(3):311-327 [FREE Full text] [doi: 10.1002/wmh3.358] [Medline: 32837779]
23. Saultz J. Essential workers. Fam Med 2020 Jun 30;52(7):469-470. [doi: 10.22454/fammed.2020.811229]
24. Ho CS, Chee CY, Ho RC. Mental health strategies to combat the psychological impact of CoronavirusDisease 2019 (COVID-19) beyond paranoia and panic. Ann Acad Med Singap 2020 Mar 31;49(3):155-160. [doi: 10.47102/annals-acadmedsg.202043]
25. Maunder R. The experience of the 2003 SARS outbreak as a traumatic stress among frontline healthcare workers in Toronto: lessons learned. Philos Trans R Soc Lond B Biol Sci 2004 Jul 29;359(1447):1117-1125 [FREE Full text] [doi: 10.1098/rstb.2004.1483] [Medline: 15306398]
26. Iacobucci G. Covid-19: all essential workers in England can now be tested. BMJ 2020 Apr 24;369:m1662. [doi: 10.1136/bmj.m1662] [Medline: 32332064]
27. White DB, Lo B. A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. JAMA 2020 May 12;323(18):1773-1774. [doi: 10.1001/jama.2020.5046] [Medline: 32219367]
28. Dennerlein JT, Burke L, Sabbath EL, Williams JAR, Peters SE, Wallace L, et al. An integrative total worker health framework for keeping workers safe and healthy during the COVID-19 pandemic. Hum Factors 2020 Aug 09;62(5):689-696 [FREE Full text] [doi: 10.1177/0018720820932699] [Medline: 32515231]
29. Farcas A, Ko J, Chan J, Malik S, Nono L, Chiampas G. Use of incident command system for disaster preparedness: a model for an emergency department COVID-19 response. Disaster Med Public Health Prep 2020 Jun 24:1-6 [FREE Full text] [doi: 10.1017/dmp.2020.210] [Medline: 32576330]
30. Lurie N, Fremont A. Building bridges between medical care and public health. JAMA 2009 Jul 01;302(1):84-86 [FREE Full text] [doi: 10.1001/jama.2009.959] [Medline: 19567444]
31. Kringos D, Carinci F, Barbazza E, Bos V, Gilmore K, Groene O, HealthPros Network. Managing COVID-19 within and across health systems: why we need performance intelligence to coordinate a global response. Health Res Policy Syst 2020 Jul 14;18(1):80 [FREE Full text] [doi: 10.1186/s12961-020-00593-x] [Medline: 32664985]
32. Digital technology for COVID-19 response. World Health Organization. 2020 Apr 03. URL: https://www.who.int/news-room/detail/03-04-2020-digital-technology-for-covid-19-response [accessed 2020-06-28]
33. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
34. Zamberg I, Windisch O, Agoritsas T, Nendaz M, Savoldelli G, Schiffer E. A mobile medical knowledge dissemination platform (HeadToToe): mixed methods study. JMIR Med Educ 2020 May 27;6(1):e17729 [FREE Full text] [doi: 10.2196/17729] [Medline: 32249758]
35. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
36. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ 2010 Sep 17;341:c4587. [doi: 10.1136/bmj.c4587] [Medline: 20851841]
37. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
38. Stanford M. Stanford Medicine COVID-19 Guide for First Responders and Essential Workers. Stanford Medicine. URL: https://med.stanford.edu/covid19/first-responder.html [accessed 2021-05-06]
39. Hendersen J, McCullough E, Treuhaft S. A profile of frontline workers in the Bay Area. Bay Area Equity Atlas. 2020 May 13. URL: https://bayareaequityatlas.org/essential-workers [accessed 2021-05-06]
40. American Community Survey 2014-2018 5-year estimates. United States Census Bureau. 2019 Dec. URL: https://www.census.gov/newsroom/press-releases/2019/acs-5-year.html [accessed 2021-05-06]
41. Hilal S, Jones D. A package deal: police, fire, and EMS all in one. The Police Chief. 2014 Sep. URL: https://www.policechiefmagazine.org/a-package-deal-police-fire-and-ems/ [accessed 2021-05-06]

XSL•FO
RenderX

42. Emergency Medical Services Fellowship. Stanford Medicine. URL: https://emed.stanford.edu/fellowships/ems.html#overview [accessed 2020-11-28]
43. San Francisco Bay Area. Bay Area Census. URL: http://www.bayareacensus.ca.gov/bayarea.htm [accessed 2020-12-16]
44. 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 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
45. Rev. URL: https://www.rev.com/ [accessed 2021-05-06]
46. Beebe J. Basic concepts and techniques of rapid appraisal. Hum Organ 1995 Mar;54(1):42-51. [doi: 10.17730/humo.54.1.k84tv883mr2756l3]
47. Brown-Johnson CG, Chan GK, Winget M, Shaw JG, Patton K, Hussain R, et al. Primary Care 2.0: design of a transformational team-based practice model to meet the quadruple aim. Am J Med Qual 2019 Nov 09;34(4):339-347. [doi: 10.1177/1062860618802365] [Medline: 30409021]
48. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
49. Miles M, Huberman A, Saldana J. Qualitative Data Analysis, 4th edition. Los Angeles, CA: SAGE Publications; 2019.
50. 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]
51. Brown-Johnson C, Safaeinili N, Zionts D, Holdsworth LM, Shaw JG, Asch SM, et al. The Stanford Lightning Report Method: a comparison of rapid qualitative synthesis results across four implementation evaluations. Learn Health Syst 2020 Apr;4(2):e10210 [FREE Full text] [doi: 10.1002/lrh2.10210] [Medline: 32313836]
52. Digregorio H, Graber JS, Saylor J, Ness M. Assessment of interprofessional collaboration before and after a simulated disaster drill experience. Nurse Educ Today 2019 Aug;79:194-197. [doi: 10.1016/j.nedt.2019.05.023] [Medline: 31154265]
53. Barnett-Vanes A, Luis Guinto RL. Disaster curricula in medical and health care education: adopting an interprofessional approach. Prehosp Disaster Med 2013 Dec 04;28(6):644-645. [doi: 10.1017/s1049023x13009114]
54. Willems A, Waxman B, Bacon AK, Smith J, Peller J, Kitto S. Interprofessional non-technical skills for surgeons in disaster response: a qualitative study of the Australian perspective. J Interprof Care 2013 Mar 25;27(2):177-183. [doi: 10.3109/13561820.2012.706337] [Medline: 22830532]
55. Sexton M, Orchard C. Understanding healthcare professionals' self-efficacy to resolve interprofessional conflict. J Interprof Care 2016 May 06;30(3):316-323. [doi: 10.3109/13561820.2016.1147021] [Medline: 27152535]
56. Hecker S. Hazard pay for COVID-19? Yes, but it's not a substitute for a living wage and enforceable worker protections. New Solut 2020 Aug 20;30(2):95-101. [doi: 10.1177/1048291120933814] [Medline: 32567480]
57. Roberts A, Nimegeer A, Farmer J, Heaney DJ. The experience of community first responders in co-producing rural health care: in the liminal gap between citizen and professional. BMC Health Serv Res 2014 Oct 18;14(1):460 [FREE Full text] [doi: 10.1186/1472-6963-14-460] [Medline: 25326796]
58. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250. [doi: 10.1016/s2468-2667(20)30086-4]
59. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
60. Bursztyn L, Rao A, Roth C, Yanagizawa-Drott D. Misinformation during a pandemic. SSRN Journal 2020. [doi: 10.2139/ssrn.3580487]
61. Vinck P, Pham PN, Bindu KK, Bedford J, Nilles EJ. Institutional trust and misinformation in the response to the 2018–19 Ebola outbreak in North Kivu, DR Congo: a population-based survey. Lancet Infect Dis 2019 May;19(5):529-536. [doi: 10.1016/s1473-3099(19)30063-5]
62. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936. [doi: 10.1056/nejmoa2001191]
63. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan 25;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]
64. Kia-Keating M, Santacrose DE, Liu SR, Adams J. Using community-based participatory research and human-centered design to address violence-related health disparities among Latino/a youth. Fam Community Health 2017;40(2):160-169 [FREE Full text] [doi: 10.1097/FCH.0000000000000145] [Medline: 28207679]
65. Durand M, Alam S, Grande SW, Elwyn G. 'Much clearer with pictures': using community-based participatory research to design and test a Picture Option Grid for underserved patients with breast cancer. BMJ Open 2016 Feb 02;6(2):e010008 [FREE Full text] [doi: 10.1136/bmjopen-2015-010008] [Medline: 26839014]
66. Gratzer D, Strudwick G, Yeung A. Mental illness: is there an app for that? Fam Syst Health 2019 Dec;37(4):336-339. [doi: 10.1037/fsh0000451] [Medline: 31815514]
67. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]

XSL•FO
RenderX

68. Lerner E, Newgard C, Mann N. Effect of the coronavirus disease 2019 (COVID-19) pandemic on the U.S. emergency medical services system: a preliminary report. Acad Emerg Med 2020 Aug;27(8):693-699 [FREE Full text] [doi: 10.1111/acem.14051] [Medline: 32557999]
69. Friedman J, Calderón-Villarreal A, Bojorquez L, Hernández C, Schriger D, Hirashima E. Excess out-of-hospital mortality and declining oxygen saturation: the sentinel role of EMS data in the COVID-19 crisis in Tijuana, Mexico. medRxiv. Preprint posted online on May 18, 2020. [doi: 10.1101/2020.05.13.20098186]

Abbreviations

EMS: emergency medical services
EMT: emergency medical technician
PPE: personal protective equipment
Q&A: question and answer
SARS: severe acute respiratory syndrome
SHC: Stanford Health Care

Edited by C Basch; submitted 18.12.20; peer-reviewed by A Garg, T Alessa, L Kremer; comments to author 01.02.21; revised version received 09.03.21; accepted 16.04.21; published 20.05.21.

Please cite as:
Vilendrer S, Amano A, Brown Johnson CG, Favet M, Safaeinili N, Villasenor J, Shaw JG, Hertelendy AJ, Asch SM, Mahoney M
An App-Based Intervention to Support First Responders and Essential Workers During the COVID-19 Pandemic: Needs Assessment and Mixed Methods Implementation Study
J Med Internet Res 2021;23(5):e26573
URL: https://www.jmir.org/2021/5/e26573
doi:10.2196/26573
PMID:33878023

©Stacie Vilendrer, Alexis Amano, Cati G Brown Johnson, Marissa Favet, Nadia Safaeinili, Jacqueline Villasenor, Jonathan G Shaw, Attila J Hertelendy, Steven M Asch, Megan Mahoney. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2021. 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 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:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31085/

Abstract

Background: More than 79.2 million confirmed COVID-19 cases and 1.7 million deaths were caused by SARS-CoV-2; the disease was named COVID-19 by the World Health Organization. Control of the COVID-19 epidemic has become a crucial issue around the globe, but there are limited studies that investigate the global trend of the COVID-19 pandemic together with each country's policy measures.

Objective: We aimed to develop an online artificial intelligence (AI) system to analyze the dynamic trend of the COVID-19 pandemic, facilitate forecasting and predictive modeling, and produce a heat map visualization of policy measures in 171 countries.

Methods: The COVID-19 Pandemic AI System (CPAIS) integrated two data sets: the data set from the Oxford COVID-19 Government Response Tracker from the Blavatnik School of Government, which is maintained by the University of Oxford, and the data set from the COVID-19 Data Repository, which was established by the Johns Hopkins University Center for Systems Science and Engineering. This study utilized four statistical and deep learning techniques for forecasting: autoregressive integrated moving average (ARIMA), feedforward neural network (FNN), multilayer perceptron (MLP) neural network, and long short-term memory (LSTM). With regard to 1-year records (ie, whole time series data), records from the last 14 days served as the validation set to evaluate the performance of the forecast, whereas earlier records served as the training set.

Results: A total of 171 countries that featured in both databases were included in the online system. The CPAIS was developed to explore variations, trends, and forecasts related to the COVID-19 pandemic across several counties. For instance, the number

XSL•FO
RenderX

of confirmed monthly cases in the United States reached a local peak in July 2020 and another peak of 6,368,591 in December 2020. A dynamic heat map with policy measures depicts changes in COVID-19 measures for each country. A total of 19 measures were embedded within the three sections presented on the website, and only 4 of the 19 measures were continuous measures related to financial support or investment. Deep learning models were used to enable COVID-19 forecasting; the performances of ARIMA, FNN, and the MLP neural network were not stable because their forecast accuracy was only better than LSTM for a few countries. LSTM demonstrated the best forecast accuracy for Canada, as the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were 2272.551, 1501.248, and 0.2723075, respectively. ARIMA (RMSE=317.53169; MAPE=0.4641688) and FNN (RMSE=181.29894; MAPE=0.2708482) demonstrated better performance for South Korea.

Conclusions: The CPAIS collects and summarizes information about the COVID-19 pandemic and offers data visualization and deep learning–based prediction. It might be a useful reference for predicting a serious outbreak or epidemic. Moreover, the system undergoes daily updates and includes the latest information on vaccination, which may change the dynamics of the pandemic.

(J Med Internet Res 2021;23(5):e27806) doi:10.2196/27806

KEYWORDS

COVID-19; artificial intelligence; time series; deep learning; machine learning; statistical analysis; pandemic; data visualization

Introduction

In December 2019, the first cases of a new respiratory disease caused by a novel coronavirus were reported in Wuhan, Hubei province, China [1]. The novel coronavirus was subsequently identified and named SARS-CoV-2, and the disease caused by SARS-CoV-2 was named COVID-19 by the World Health Organization (WHO) [2,3]. Since the time the first cases were reported, many confirmed cases have been reported in various other countries. By March 11, 2020, more than 118,000 confirmed cases and 4291 deaths had been reported across 114 countries. The WHO declared the COVID-19 outbreak a pandemic [4], which continues to worsen. As of December 27, 2020, there were more than 79.2 million confirmed cases and 1.7 million deaths [5]. COVID-19 management has emerged as an urgent global issue. Many studies have investigated the factors that contribute to the spread of COVID-19. Demographic, geographic, and economic factors have influenced the spread of the disease. However, social factors, especially governmental response to the pandemic, have significantly influenced disease severity within certain countries [6-11]. Some countries have shown that implementing rigorous public health care management strategies can successfully control infection spread and maintain normal societal functioning [11].

The rapid development of artificial intelligence (AI) in the health care field offers new opportunities to medical researchers. There are many studies that employ AI techniques in disease predictions, such as Yu et al, who have established an online machine learning health assessment system for metabolic syndrome and chronic kidney diseases [12]. Lin et al utilized multicenter data to develop an end-stage liver disease mortality prediction scoring system [13]. Ayyoubzadeh et al analyzed the rate of COVID-19 incidence in Iran using Google Trends data and deep learning methods [14]. Yeung et al combined several online COVID-19 data to train and evaluate five non–time series machine learning models in predicting confirmed infection growth [15]. These studies have shown that AI is suitable for evaluating disease trends and can provide governments with information that can be used to prevent outspread. There are abundant research findings on COVID-19–related AI prediction and the utilization of mobile sensor data with cell broadcast to identify and manage potential contacts [14,16-20].

However, most of these studies have been conducted in a specific region or single country. There is public health consensus that vaccination is an effective prevention strategy. However, with regard to its efficiency and medical expenditure, long-term follow-up investigation is needed to evaluate the clinical effects of vaccines that have not undergone the standard approval process and tests of their mid- and long-term side effects on different groups [21]. Moreover, different studies have focused on different time frames in pandemic trend prediction. They have drawn the same conclusion: there is a high possibility that COVID-19 will remain a common illness or become endemic in the future, and we must learn to coexist with it. Many factors influence how the pandemic will progress (eg, herd immunity), and governmental and individual responses vary widely across nations [22,23]. Successful epidemic prevention and control measures remain the most efficient solution for public health problems. However, there is limited literature on the relationship between governmental responses and the severity of the domestic spread of COVID-19 [24,25].

Therefore, we constructed an online AI system that contains worldwide COVID-19–related data, each country's governmental responses to the COVID-19 pandemic, and each country's population data [26]. The COVID-19 Pandemic AI System (CPAIS) can be used to analyze the dynamic trend of the COVID-19 pandemic, facilitate forecasting and predictive modeling, and produce heat map visualization of policy measures in different countries.

Methods

Data Acquisition and System

The CPAIS integrated two data sets: the data set from the Oxford COVID-19 Government Response Tracker (OxCGRT) from the Blavatnik School of Government, which is maintained by the University of Oxford, and the data set from the COVID-19 Data Repository, which was established by Johns Hopkins

XSL•FO
RenderX

University Center for Systems Science and Engineering (CSSE). The COVID-19 Data Repository also contains each country's population data, which are obtained from the United Nations World Population Prospects [27-31]. A total of 171 countries that featured in the databases were included in the system.

The CPAIS was placed on a sever and embedded with time series deep learning models to provide forecasting analyses by the statistical program R, version 3.6.3 (The R Foundation). We used the React.js, version 16.14.0, framework; the styling language Sass (Syntactically Awesome Style Sheets), version 4; and the programming language JavaScript ES6 for front-end implementation. As for back-end implementation, we used Java 8; Spring Boot, version 2.0.2 (VMware, Inc); and R as the programming languages, and we used the MySQL (Structured Query Language), version 5.7.21, database as the storage system. In addition, this AI-based system has been programmed to update itself by auto-retrieving information from all data sets each morning at 9 AM (GMT + 8). The auto-retrieval can be summarized in the following three steps: (1) setting the crawler to fetch the data from the source databases, (2) integrating the updated data into our own MySQL database, and (3) conducting statistical analysis using the database-stored procedure.

The COVID-19 Data Repository established by Johns Hopkins University CSSE contains three categories of data concerning COVID-19 incidence—confirmed cases, recovered cases, and number of deaths—with country geolocation retrieved from 192 affected countries since January 21, 2020. For most of the countries, country-level data concerning the numbers of reported cases are available. Province- and city-level data concerning reported cases are available for some countries. To depict the COVID-19 pandemic comprehensively, we archived country-level data. The number of reported cases was updated daily using data retrieved from multiple online sources. The number of cases was retrieved from the WHO and the regional and local health departments of the affected countries, including their centers for disease control and prevention. All data were shared freely through GitHub.

OxCGRT has been collecting and documenting governmental responses to the COVID-19 pandemic based on several parameters since January 1, 2020. The data set includes 183 countries and 20 items (19 indicators and 1 free response) that characterize governmental responses. There are three types of items: (1) ordinal scale for severity or intensity, (2) numeric scale for specific numbers, and (3) text for other information types. These items can be further classified into four groups: (1) containment and closure policies (8 indicators), (2) economic policies (4 indicators), (3) health system policies (7 indicators), and (4) miscellaneous policies (1 free response). Miscellaneous policies were not included in this system because they were assessed using a free-text response format and limited data were available. OxCGRT data were retrieved from publicly available sources and regularly updated on GitHub.

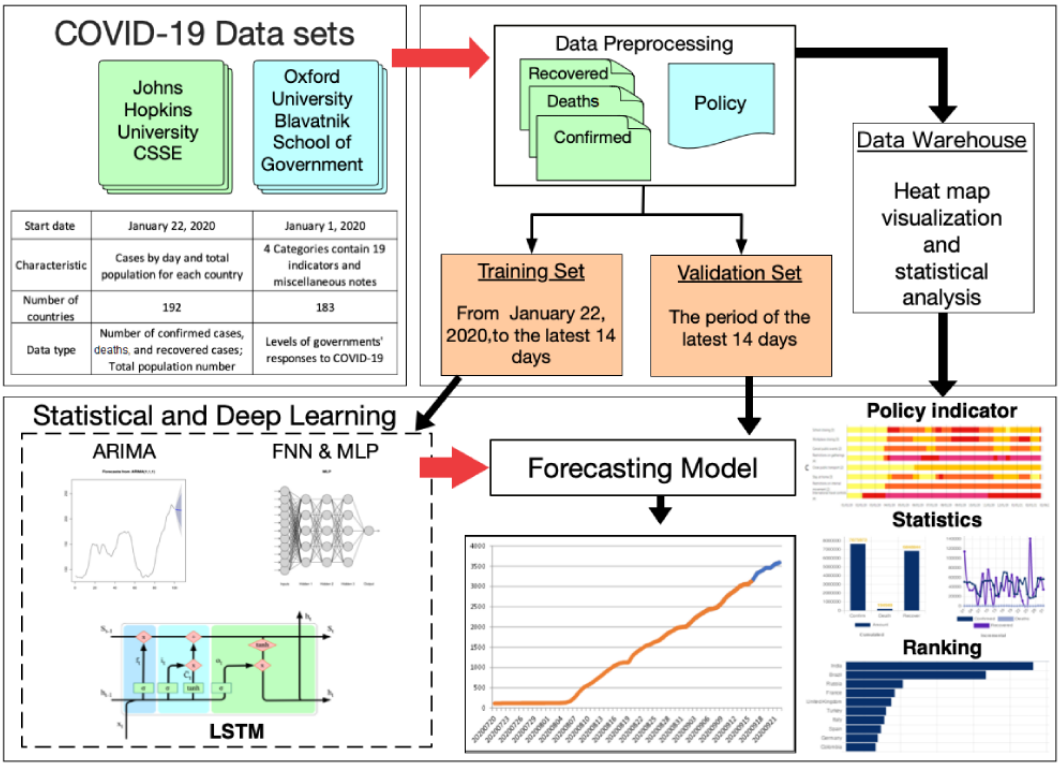
Statistical Analysis and Deep Learning Techniques

Overview

Four time series models were considered for this study. Each model was applied to all the countries in our system to facilitate forecasting. With regard to 1-year records (ie, whole time series data), records from the last 14 days served as the validation set, whereas earlier records served as the training set. Using records from the last 14 days, forecasting performance was evaluated based on the following five indices: mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE) [32,33]. RMSE, MAE, and MAPE are always positive values, whereas RMSE, MPE, and MAPE are scaled measures. The hyperparameters for each model can be found in Table S1 in Multimedia Appendix 1, and the diagram of the neural networks can be found in Figure 1. R, version 3.6.3 (The R Foundation), was used to conduct statistical analysis and apply deep learning techniques.

XSL•FO
RenderX

Figure 1. The structure of the COVID-19 Pandemic AI System (CPAIS). ARIMA: autoregressive integrated moving average; CSSE: Center for Systems Science and Engineering; FNN: feedforward neural network; LSTM: long short-term memory; MLP: multilayer perceptron; NN: neural network.



Autoregressive Integrated Moving Average

An autoregressive integrated moving average (ARIMA) model is a statistical regression analysis that utilizes time series data to either understand the data set better or predict future trends. The purpose of ARIMA is to forecast future trends by examining differences between values in the series rather than by using actual values [34,35]. The three main components of ARIMA are autoregression, integration, and moving average. Autoregression refers to a model with a changing variable that regresses on its lag values. Integration represents the differences between data values and their previous values for stationary time series. Moving average incorporates the dependence between an observation and an error term from a moving average model. An ARIMA model can be comprehended by outlining each component, which serves as a parameter with a standard notation. For ARIMA models, there are three standard notations, wherein integer values serve as substitutes for the parameters to indicate the type of ARIMA model used.

The parameters can be defined as follows:

- p: the number of time lags
- d: the degree of differencing
- q: the size of the moving average window.

In this study, we used the *auto.arima* function for R, which returns the best ARIMA model based on either the Akaike information criterion value or Bayesian information criterion value. The function searches for possible models within the order constraints provided in the *forecast* package for R [36,37].

Feedforward Neural Network

A feedforward neural network (FNN) is the simplest type of artificial neural network [38]. The FNN algorithm is biologically inspired. It consists of several simple neuron-like units that are organized in layers. In FNN, information moves in one direction—from the input nodes, through the hidden nodes, and to the output nodes. The mechanism of an FNN is different from that of recurrent neural networks (RNNs) in that connections between the units do not form cycles or loops in FNNs [38,39]. In this study, we used the *nnetar* function for R, which constructs FNNs with a single hidden layer and lagged inputs for the purpose of forecasting univariate time series. Also, in the *forecast* package, the function fits into a single hidden-layer neural network for forecasting, with the *nnet* function included in the *nnet* package for R [40,41].

Multilayer Perceptron Neural Network

Like FNNs, multilayer perceptron (MLP) neural networks are common deep learning feedforward networks. An MLP neural network is also a supervised learning algorithm used for classification. The main difference is that between the input and output layer, there can be multiple nonlinear layers, called hidden layers, which are the true computational engine of the MLP neural network. MLP neural networks use a learning technique called back-propagation for training. Their multiple layers and nonlinear activation distinguish MLP neural networks from a linear perceptron [42-44]. In other words, MLP neural networks are designed to solve nonlinearly separable problems. Specifically, the units of MLP neural networks apply a sigmoid function as an activation function. In the back-propagation technique, the difference between the output values and the *ground truth* answer are calculated using predefined error

XSL•FO
RenderX

functions. The error is fed back through the network. Using this information, the algorithm can adjust the weights of each connection to significantly reduce the value of the error function. In this study, the *mlp* function fits MLP neural networks for time series forecasting executed using the *nnfor* package [45-47].

Long Short-term Memory

Long short-term memory (LSTM) networks are a special type of recurrent deep learning neural network that learns order dependence in sequence prediction problems. LSTM was introduced by Hochreiter and Schmidhuber in 1997, and it is now widely used in a variety of studies and projects [48,49]. A typical RNN makes use of sequential information. These networks are described as *recurrent* because they use their internal state to process the variable length sequences of inputs. It is difficult for a standard RNN to carry forward information from prior time steps to later ones if a sequence is too long, because it may exclude important information from the beginning. Therefore, LSTM has an advantage in that information can be remembered for long periods of time. Unlike traditional FNNs, LSTM has feedback connections, whereby the output from the previous step is supplied as input in the current step [50]. A common LSTM unit includes a cell, an input gate, an output gate, and a forget gate. The cell recalls values over an arbitrary time interval, and the three gates regulate the flow of information in and out of the cell. In this study, we used the *keras* R package to recall TensorFlow for conducting the LSTM analysis [51]. TensorFlow was developed by the Google Brain team and released in 2015. It is a free open-source software library for machine learning techniques, particularly deep neural networks [52].

Data Visualization of Time Series Data Sets

Heat maps can be generated to depict variations in policy measures for the COVID-19 pandemic across time. Gradient color bars represent changes in measures across different levels and the support received in the form of financial assistance and investments. The time schedule presented along the horizontal axis will be updated daily. Cumulative and monthly records are represented using histograms and line charts, respectively. This system also provides a download option to interested countries and comparable services with dynamic rankings of the total number of confirmed cases and deaths and declining trends for the COVID-19 pandemic. The following simple regression formula is used to examine declining trends with dynamic time intervals:

$$y_i = \alpha + \beta x_i$$

where β is the slope that represents an increasing or decreasing trend.

Results

In this study, the CPAIS was developed to explore variations, trends, and forecasts related to the COVID-19 pandemic across several counties. A drop-down list for country selection is available. The framework of the CPAIS—from data acquisition and preprocessing to deep learning model application, forecasting, and data visualization—is presented in Figure 1. It includes a combination of two data sets, construction of databases for deep learning prediction and statistical analysis, four statistical or deep learning models for forecasting, and front-end functions for data visualization.

The numbers of confirmed cases, recovered individuals, and deaths in 15 countries are listed by month in Table 1. The number of confirmed monthly cases in the United States reached a local peak in July 2020 and another peak of 6,368,591 in December 2020. Regarding the United States, the number of recovered cases after December 14, 2020, is not recorded in the COVID-19 Data Repository database. The total population for each of the 15 countries in 2020 is also mentioned in the table. The dynamic heat map with policy measures is shown in Figure 2, which depicts changes in COVID-19 measures for each country, with Australia used as an example. A total of 19 measures were embedded within the three main policy sections (ie, containment and closure policies, economic policies, and health system policies). Economic policies have the least number of measures, and only 4 of the 19 measures are continuous measures related to financial support or investment.

Deep learning and statistical learning models were used to enable COVID-19 forecasting. The function facilitates 14-day forecasting using four powerful algorithms (Figure 3). ARIMA is the statistical learning model with time series regression; the other models are deep learning neural network algorithms with a single hidden layer, multiple hidden layers, or recurrent techniques. The performance of forecasting for each model for the 15 countries listed in Table 1 is shown in Table 2. A small error value indicates a perfect fit for the data, but the comparison between the different countries was not meaningful because they had different baselines based on their populations. For most of the countries, LSTM demonstrated better forecast accuracy with fewer errors than the other models. The performances of ARIMA, FNN, and the MLP neural network were not stable because their forecast accuracy was only competitive with LSTM for some specific countries. For example, LSTM demonstrated the best forecast accuracy for Canada. The RMSE, MAE, and MAPE were 2272.551, 1501.248, and 0.2723075, respectively. ARIMA (RMSE=317.53169; MAPE=0.4641688) and FNN (RMSE=181.29894; MAPE=0.2708482) demonstrated better performance for South Korea.

Figure 4 presents descriptive statistics for specific countries. On the website, three countries can be simultaneously compared, and the period can be customized. Users can select the countries that are of interest to them and compare the COVID-19–related data. For each respective country, a line chart showing the number of confirmed cases, recoveries, and deaths per month is generated. In addition, a global comparison is also provided on the website.

Users can rank 171 countries based on five different parameters: (1) the number of confirmed cases, (2) confirmed cases by percentage of population, (3) the number of confirmed deaths, (4) confirmed deaths by percentage of population, and (5) declining trend. Figure 5 shows an example of how the top 20 countries can be ranked using confirmed cases by percentage of population. With regard to customization, the ranking

XSL•FO
RenderX

function is flexible. The selected countries and specific time period can be changed by the user.

XSL•FO
RenderX

Table 1. The numbers of confirmed cases, recovered individuals, and deaths in 15 countries by month in 2020.

Country (total population[a]) and cases	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
United States (N=329,466,283)												
Confirmed	7	17	192,152	884,047	718,241	834,359	1,922,730	1,464,676	1,201,822	1,914,993	4,466,451	6,368,591
Deaths	0	1	5271	60,699	41,703	20,113	26,306	29,591	23,515	23,928	37,038	77,572
Recovered	0	7	7017	146,923	290,811	275,873	717,529	746,665	655,863	771,790	1,533,841	1,151,763[b]
Canada (N=37,855,702)												
Confirmed	4	16	8507	45,930	38,022	13,618	12,184	12,637	30,189	76,206	144,244	202,852
Deaths	0	0	101	3209	4064	1276	330	193	173	841	1960	3485
Recovered	0	0	1586	19,832	27,789	19,907	33,786	13,114	21,161	61,771	105,643	189,043
Mexico (N=127,792,286)												
Confirmed	0	4	1211	18,009	71,440	135,425	198,548	174,923	143,656	181,746	181,746	312,551
Deaths	0	0	29	1830	8071	17,839	18,919	17,726	13,232	14,107	14,107	19,867
Recovered	0	0	35	11,388	52,349	110,766	152,577	169,107	131,785	151,364	151,364	251,209
Brazil (N=212,559,409)												
Confirmed	0	2	5715	81,470	81,470	887,192	1,260,444	1,245,787	902,663	724,670	800,273	1,340,095
Deaths	0	0	201	5805	5805	30,280	32,881	28,906	22,571	15,932	13,236	21,829
Recovered	0	0	127	35,808	35,808	581,763	1,220,536	1,259,737	1,006,183	730,387	592,641	1,251,042
Argentina (N=45,195,777)												
Confirmed	0	0	1054	3374	12,423	47,679	126,772	226,433	333,266	415,923	257,609	200,981
Deaths	0	0	27	191	321	768	2236	5117	8277	14,065	7728	4515
Recovered	0	0	240	1016	4080	16,692	61,752	217,415	293,450	379,294	283,288	169,449
Chile (N=19,116,209)												
Confirmed	0	2	2842	14,858	105,848	155,843	76,274	56,059	51,265	47,265	41,487	57,230
Deaths	0	0	12	215	827	4634	3769	1832	1452	1466	1203	1198
Recovered	0	0	156	8424	34,147	198,502	87,098	55,552	52,710	50,053	39,962	50,778
United Kingdom (N=67,886,004)												
Confirmed	2	59	38,754	139,956	78,768	27,677	19,577	33,290	117,763	558,947	618,940	862,498
Deaths	0	0	2457	24,297	10,773	2952	795	315	644	4412	11,900	15,077
Recovered	0	8	171	680	331	180	69	243	691	466	731	1909
France (N=65,273,512)												
Confirmed	5	95	52,727	114,472	21,710	13,054	23,134	93,789	285,045	808,678	864,165	400,792
Deaths	0	2	3530	20,847	4426	1041	422	372	1346	4840	15,993	11,940
Recovered	0	12	9501	39,963	18,997	7926	5365	5026	11,842	24,463	44,818	32,229
Greece (N=10,423,056)												
Confirmed	0	4	1310	1277	326	492	1068	5840	8158	20,776	66,020	33,579
Deaths	0	0	49	91	35	17	14	60	125	235	1780	2432
Recovered	0	0	52	1322	0	0	0	2430	7882	11,388	0	70,690
Taiwan (N=23,816,775)												
Confirmed	9	29	283	107	13	5	20	21	26	41	120	124
Deaths	0	1	4	1	1	0	0	0	0	0	0	0
Recovered	0	9	30	283	101	14	3	22	21	32	50	106

XSL•FO
RenderX

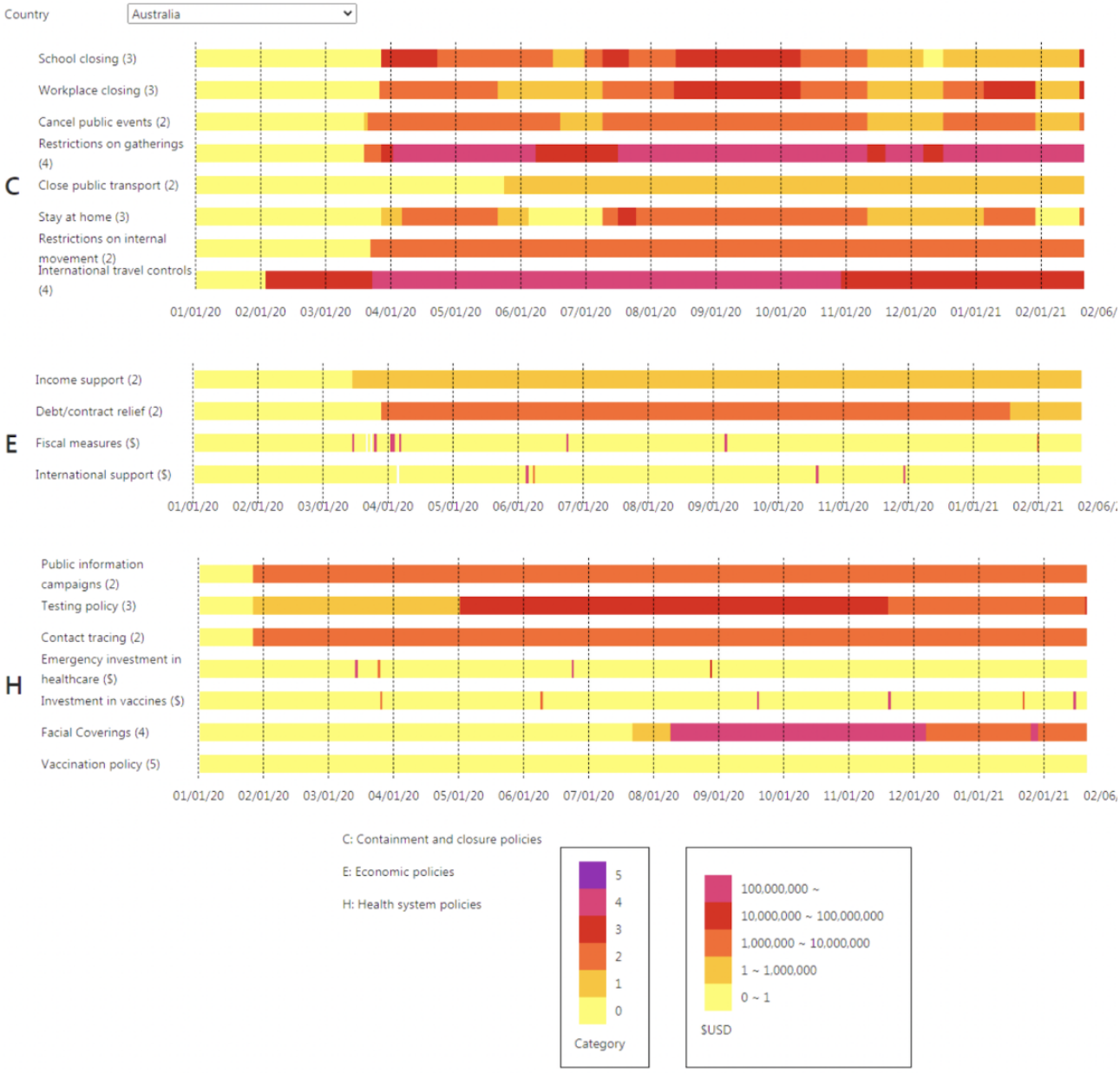
Country (total population[a]) and cases	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Thailand (N=69,799,978)												
Confirmed	17	23	1609	1303	127	90	139	107	152	215	224	3155
Deaths	0	0	10	44	3	1	0	0	1	0	1	3
Recovered	5	23	314	2342	279	93	69	149	105	213	219	462
South Korea (N=51,269,183)												
Confirmed	10	313	6636	988	729	1347	1486	5846	3707	2746	8017	27,117
Deaths	0	16	146	86	23	11	19	23	91	51	60	391
Recovered	0	27	5381	3664	1350	1191	1620	1965	6468	2691	3528	15,068
India (N=1,380,004,385)												
Confirmed	1	2	1394	33,466	155,746	394,872	1,110,507	1,995,178	2,621,418	1,871,498	1,278,727	803,865
Deaths	0	0	35	1119	4254	11,992	19,111	28,777	33,390	23,433	15,510	11,117
Recovered	0	3	120	8945	82,784	256,060	746,462	1,745,508	2,433,319	2,218,312	1,398,072	970,695
Australia (N=25,459,700)												
Confirmed	9	16	4534	2207	436	718	9360	8539	1277	499	317	513
Deaths	0	0	18	75	10	1	97	456	231	19	1	1
Recovered	2	9	347	5384	876	422	2943	11,367	3434	552	266	160
Egypt (N=102,334,403)												
Confirmed	0	1	709	4827	4827	43,326	25,767	4861	4259	4357	8356	22,151
Deaths	0	0	46	346	346	1994	1852	616	509	336	384	981
Recovered	0	1	156	1224	1224	12,423	21,178	33,291	23,565	2958	3266	9387

[a]Total population in 2020.

[b]The number of recovered cases after December 14, 2020, were not recorded in the COVID-19 Data Repository database (the record only includes cases from December 1 to 14, 2020); therefore, this value was underreported.

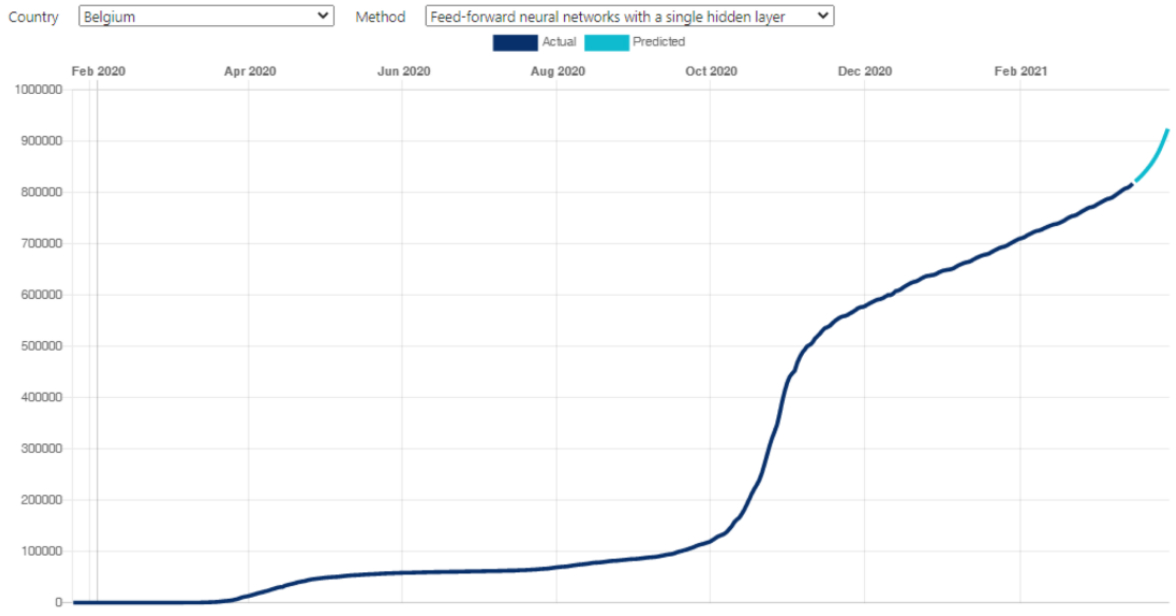
XSL•FO
RenderX

Figure 2. The interface of the dynamic heat map with policy measures on the COVID-19 Pandemic AI System (CPAIS) website.



XSL•FO
RenderX

Figure 3. The COVID-19 Pandemic AI System (CPAIS) interface for machine learning prediction models facilitating 14-day COVID-19 forecasting. The plot shows the curve for deep learning modeling of total cumulative confirmed cases.



XSL•FO
RenderX

Table 2. Forecasting performance for each model in the validation set for the 15 countries.

Country (total population[a]) and methods	Mean error[b]	Root mean square error[b]	Mean absolute error[b]	Mean percentage error[b]	Mean absolute percentage error[b]
United States (N=329,466,283)					
ARIMA[c]	–183,472.5153	229,501.345	183,888.691	–0.9538265	0.9562102
FNN[d]	–197,967.69975	251,014.19	201,574.807	–1.027988	1.048648
MLP[e]	34,016.71589	45,932.609	35,569.561	0.1774821	0.1862749
LSTM[f]	–17,670.38	*41,667.98* [g]	*31,092.06*	–0.09409045	*0.1664009*
Canada (N=37,855,702)					
ARIMA	–3786.81463	4953.7659	3786.8146	–0.6828342	0.6828342
FNN	–1902.8218773	3146.8161	2133.5721	–0.3503041	0.3898707
MLP	–6056.7104430	7294.1933	6056.7104	–1.094643	1.094643
LSTM	306.1702	*2272.551*	*1501.248*	0.04896196	*0.2723075*
Mexico (N=127,792,286)					
ARIMA	–3776.6237	6281.987	4841.2544	0.3501243	1.2391347
FNN	–15,894.200241	19,622.066	16,156.1290	–1.145524	1.165534
MLP	–3551.381635	6534.119	5455.281	–0.2517612	0.3969063
LSTM	–1137.118	*2883.836*	*2334.178*	–0.08386455	*0.1716616*
Brazil (N=212,559,409)					
ARIMA	–52,913.8661	69,053.95	54,328.55	–0.7032164	0.7228866
FNN	–168,251.54394	204,577.061	168,251.544	–2.240681	2.240681
MLP	–28,723.33938	43,395.965	31,117.856	–0.3797225	0.412664
LSTM	–2746.457	*16,085.02*	*14,347.73*	–0.03768765	*0.1931052*
Argentina (N=45,195,777)					
ARIMA	10,240.495912	12,832.6035	10,240.4959	0.6433934	0.6433934
FNN	22,285.962404	26,555.128	22,285.9624	1.402042	1.402042
MLP	10,914.143275	13,689.5539	10,929.6874	0.6857769	0.6867919
LSTM	1253.045	*3920.961*	*3202.607*	0.07803485	*0.2024643*
Chile (N=19,116,209)					
ARIMA	1823.55216	1992.35	1823.5522	0.3048502	0.3048502
FNN	8171.7723060	9157.9881	8171.7723	1.363951	1.363951
MLP	2169.702307	2435.4540	2169.7023	0.3622628	0.3622628
LSTM	595.9308	*790.8397*	*648.5224*	0.1001373	*0.1090634*
United Kingdom (N=67,886,004)					
ARIMA	40,161.7481	55,436.735	41,580.2155	1.7053944	1.776331
FNN	–17,129.950943	23,936.144	17,129.951	–0.7304511	0.7304511
MLP	81,031.84	102,155.3238	81,031.841	3.482155	3.482155
LSTM	15,560.98	*17,735.29*	*15,560.98*	0.6832804	*0.6832804*
France (N=65,273,512)					
ARIMA	1807.5070	*8181.384*	*6633.665*	0.07287266	*0.2565254*
FNN	61,075.99023	67,684.575	61,075.990	2.340844	2.340844
MLP	9601.594851	11,456.382	10,239.308	0.3726648	0.3969022
LSTM	6262.693	9254.264	7784.804	0.241549	0.3000627

XSL•FO
RenderX

Country (total population[a]) and methods	Mean error[b]	Root mean square error[b]	Mean absolute error[b]	Mean percentage error[b]	Mean absolute percentage error[b]
Greece (N=10,423,056)					
ARIMA	5423.2143	6072.0773	5423.2143	4.003338	4.003338
FNN	–21.8694361	*561.98452*	*400.61927*	–0.01977488	*0.2937978*
MLP	–1145.165405	1341.1596	1145.1654	–0.844399	0.844399
LSTM	–512.1191	565.7909	512.1191	–0.3821559	0.3821559
Taiwan (N=23,816,775)					
ARIMA	–15.97434477	17.288501	15.97434	–2.0379969	2.037997
FNN	–6.571007146	7.379679	6.571007	–0.84606232	0.8460623
MLP	–9.485179	12.925238	9.9162023	–1.2005706	1.257011
LSTM	–2.059649	*3.322996*	*2.978151*	–0.3227033	*0.3820354*
Thailand (N=69,799,978)					
ARIMA	1471.082153	1620.87009	1471.082153	23.7842238	23.784224
FNN	1463.109910	1611.239573	1463.109910	23.659524	23.659524
MLP	1517.21984066	1674.585004	1517.219841	24.5165025	24.516502
LSTM	173.2286	*308.695*	*202.2714*	2.950519	*3.435209*
South Korea (N=51,269,183)					
ARIMA	–260.265311	317.53169	265.29603	–0.4540395	0.4641688
FNN	–75.7162332	*181.29894*	*154.2065*	–0.1226205	*0.2708482*
MLP	–1138.0352476	1419.83911	1145.57606	–1.963196	1.978379
LSTM	323.9709	342.9156	323.9709	0.5978793	0.5978793
India (N=1,380,004,385)					
ARIMA	19,113.77834	21,947.375	19,113.778	0.1874688	0.1874688
FNN	–10,156.962689	*13,612.018*	*10,156.963*	–0.09945817	*0.09948717*
MLP	20,964.3576266	24,556.936	20,964.358	0.2055718	0.20055718
LSTM	–13,037.64	14,480.91	13,037.64	–0.128178	0.1281378
Australia (N=25,459,700)					
ARIMA	26.9606020	30.40208	26.96060	0.09542063	0.09542063
FNN	187.8959192	205.6998	187.89592	0.6634038	0.6637038
MLP	–15.69085695	76.48186	62.261210	–0.05478576	0.2197826
LSTM	5.898776	*14.39023*	*11.91991*	0.02086999	*0.04212132*
Egypt (N=102,334,403)					
ARIMA	2392.285714	3239.04732	2392.28571	1.7844594	1.784459
FNN	1944.5586880	2641.98168	1944.55869	1.45017	1.45017
MLP	669.96030638	936.05245	669.96031	0.4988667	0.4988667

XSL•FO
RenderX

Country (total population[a]) and methods	Mean error[b]	Root mean square error[b]	Mean absolute error[b]	Mean percentage error[b]	Mean absolute percentage error[b]
LSTM	437.0412	*500.6487*	*438.0092*	0.3304228	*0.3311979*

[a]Total population in 2020.

[b]Five commonly used measures for evaluation of forecasting include mean error, root mean square error (RMSE), mean absolute error (MAE), mean percentage error, and mean absolute percentage error (MAPE), according to the records of the latest 14 days in 2020. The RMSE, MAE, and MAPE are always positive values.

[c]ARIMA: autoregressive integrated moving average.

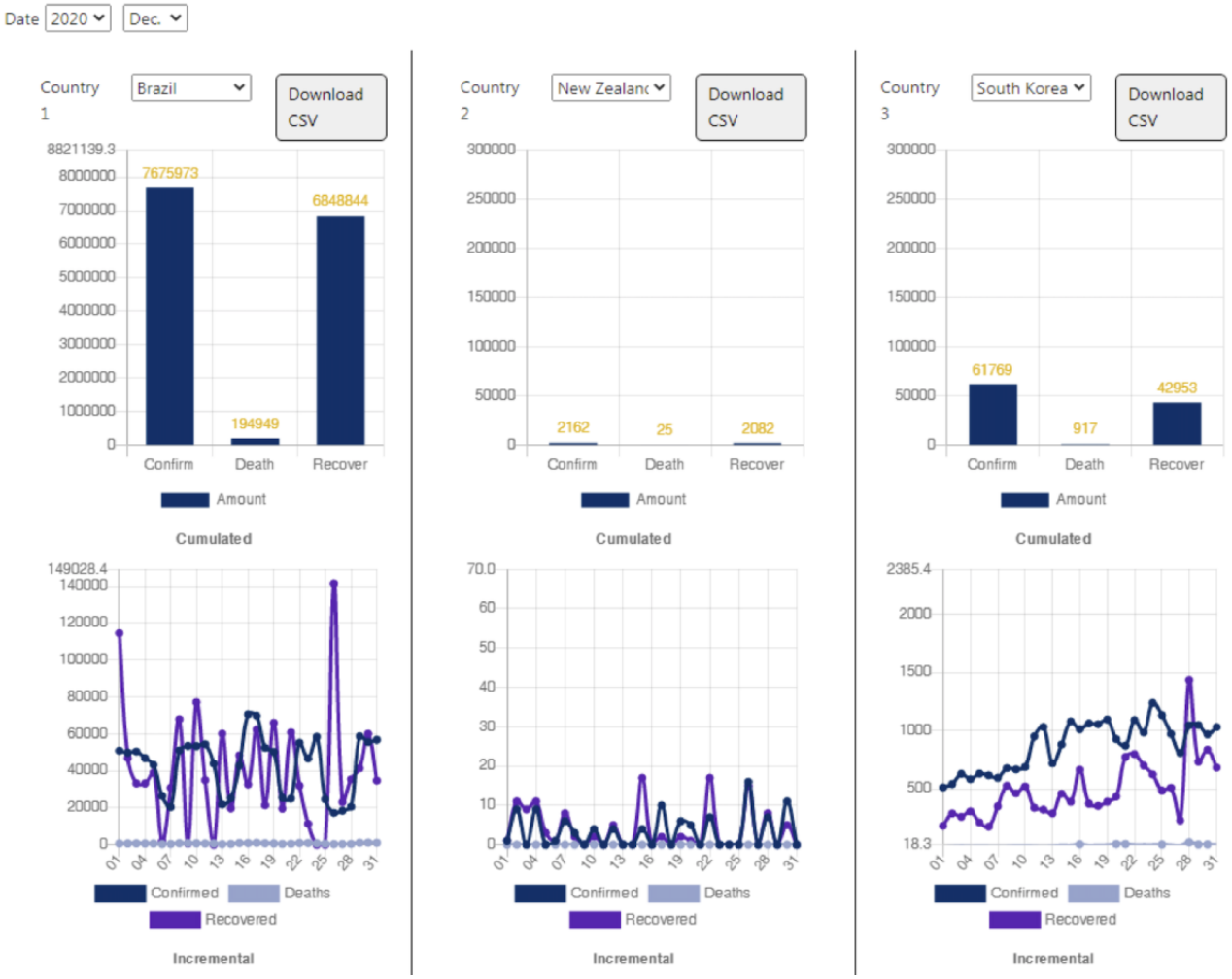
[d]FNN: feedforward neural network.

[e]MLP: multilayer perceptron.

[f]LSTM: long short-term memory.

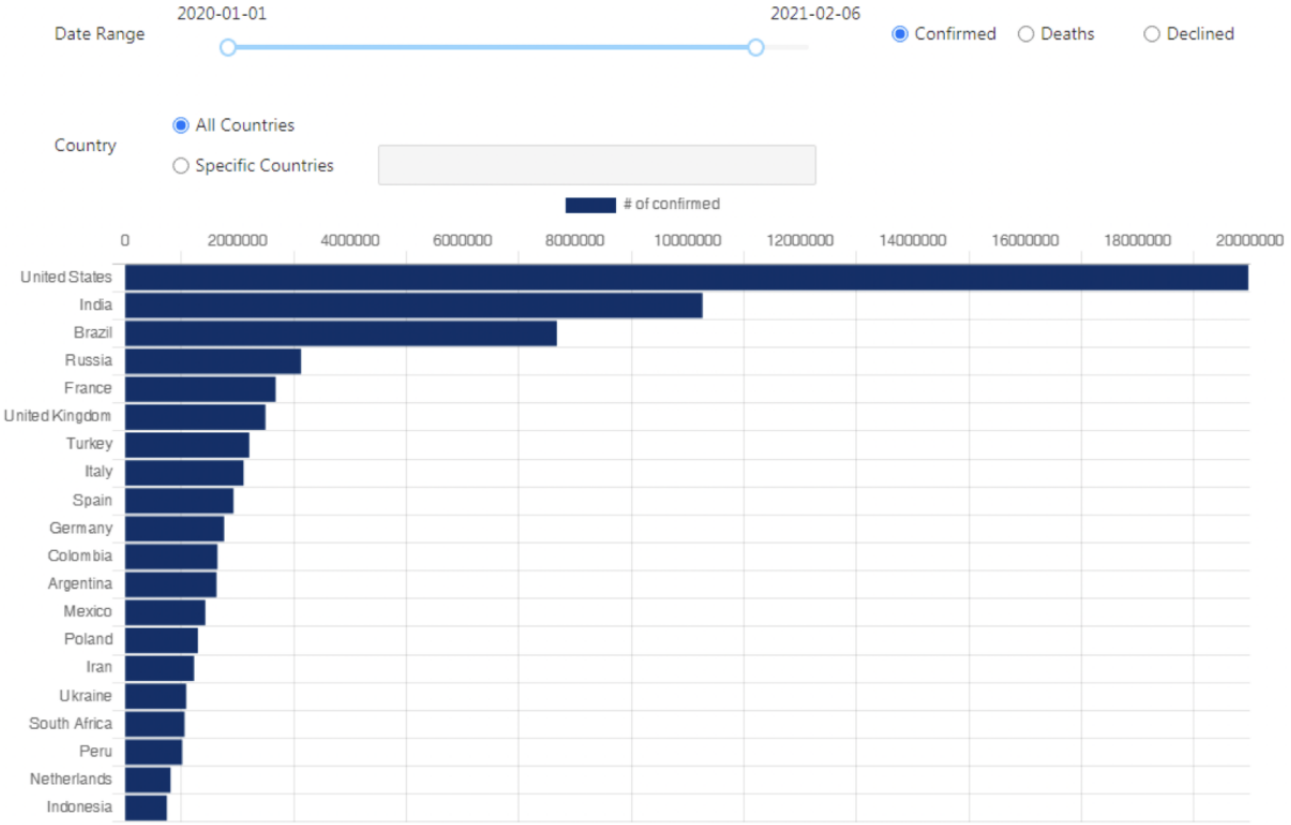
[g]The values for best performances in each country are italicized.

Figure 4. The interface of descriptive statistics for selected countries with customization on the COVID-19 Pandemic AI System (CPAIS) website. CSV: comma-separated values.



XSL•FO
RenderX

Figure 5. The interface for the ranking of selected countries with customization on the COVID-19 Pandemic AI System (CPAIS) website.



Discussion

Principal Findings

A combination of data on COVID-19 incidence and policy measures can be used to examine the relationship between the progression of the COVID-19 pandemic and governmental epidemic prevention efforts. The CPAIS can help users determine whether policy measures are successful in preventing COVID-19 transmission. According to a report published by the Lowy Institute for International Policy [53], a ranked comparison of the performance of countries in managing the COVID-19 pandemic shows that New Zealand, Vietnam, and Taiwan are the top three countries with the highest average scores on their six indicators. Besides, New Zealand and Taiwan successfully controlled the COVID-19 outbreak without international financial support (Figures S1-S3 in Multimedia Appendix 1). Specifically, New Zealand had immediately implemented infection control and closure policies with a flexible adaptation on measures; in addition, Taiwan had enforced strict guidelines regarding international travel that not only contributed to infection control but also rendered the strict measures described in the containment and closure policies unnecessary. Furthermore, both countries had taken great efforts to maximize the implementation of testing and contact tracing policies during 2020. In this regard, both countries are outstanding examples. The vivid heat maps in the CPAIS illustrate time-dependent fluctuations in the measures and help users monitor variations in, and the effects of, policy measures in each country.

Several time series AI learning techniques have been used for forecasting purposes. Both statistical learning and deep learning models demonstrated efficacious performance for different countries. Although the values are not absolute, they are comparable between countries with different total populations. When compared to the results of a past study [19], performance for the same model and country was better in this study because more extensive time series data were included in our system. In addition, 14-day COVID-19 trend forecasting can serve a useful alert that will help governments and experts reduce the incidence of COVID-19. Furthermore, different AI learning techniques have unique advantages.

According to the Wold decomposition theorem [34,54,55], the autoregressive moving average model is theoretically sufficient to describe a regular stationary time series. It is possible to change a nonstationary time series into a stationary one, such as by using differencing. As noted earlier, ARIMA models have three components: autoregression, integration, and moving average. They are applied to data with evidence of nonstationarity in the mean, whereby an initial differencing step can be applied one or more times to eliminate the nonstationarity of the mean function in the trend. We used the *auto.arima* function for R to choose the best model according to either the Akaike information criterion, corrected Akaike information criterion, or Bayesian information criterion value; the *auto.arima* function also conducts the model search within the order constraints provided. FNN is similar to ARIMA because the fitted model is analogous to an autoregression(p) model, where p is the order but with nonlinear functions for nonseasonal data in this study. Therefore, it is denoted as a neural network autoregression(p,k) model called *NNAR*, where k represents the

XSL•FO
RenderX

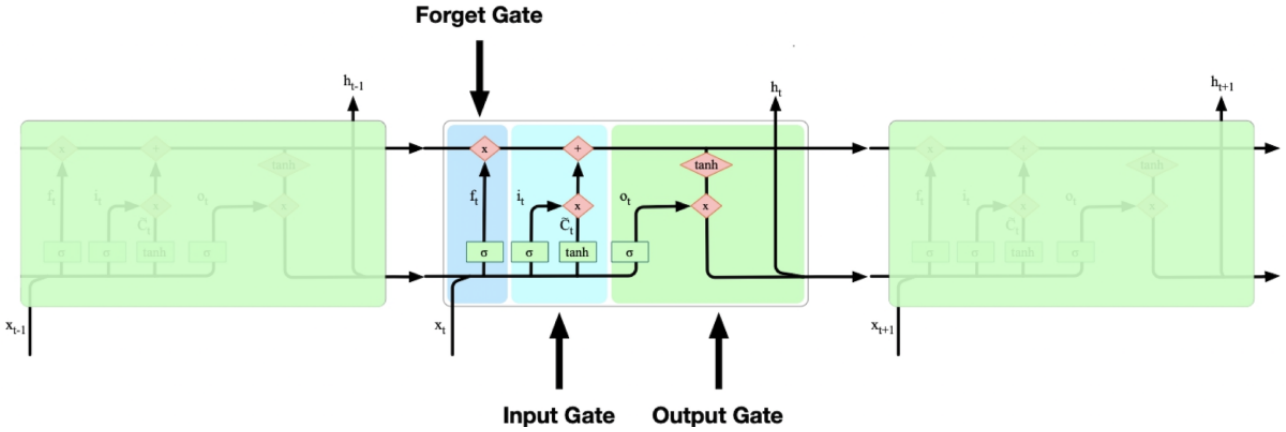
number of hidden nodes. That is why, for some countries, ARIMA and FNN yielded similar outcomes for forecast accuracy. Differences between the two models still exist; the error can be reduced only for FNN by increasing the number of iterations, but the iteration time will be increased as a result.

The capabilities of neural networks are attributable to the hierarchical or multilayered structure of the networks. The data structure can include features at different scales or resolutions and combine them into higher-order features. After repeating the learning process for a sufficient number of training cycles, the network will transition to some state where the error term is small enough. Generalization and tolerance are the two main characteristics. First, neural networks permit generalization because they can classify both unknown and known patterns with the same distinguishing features. Second, neural networks are highly fault tolerant. Because of their distributed nature, they will continue to function even if a significant fraction of neurons and interconnections fail. In general, increasing the number of hidden nodes may enhance the performance of prediction, and increasing the number of networks to train may result in an ensemble forecast.

The core idea of LSTM lies in the cell state—the horizontal line that runs down the chain with information flowing alongside (Figure 6). In addition, LSTMs have the ability to remove or add information to the cell state, controlled by the gates, which are a pathway through which information can be allowed to pass. They consist of a sigmoid neural net layer and a pointwise multiplication operation. LSTM networks are powerful in promptly forecasting series data since there can be lags of unknown duration between events in time series. Hence, when compared to other traditional RNNs in this study, LSTM networks do not have the vanishing gradient problem. Thus, LSTM has the advantages of being relatively insensitive to time intervals and of making fewer errors in prediction when compared to other methods.

In the CPAIS, long-term cumulative records of confirmed cases, recoveries, and deaths are included. In addition, daily figures for these metrics are provided for each month. Thus, short-term trends can be examined using this system. Users can compare three or more countries and visualize the relative incidence of COVID-19 within a specific time duration. Short-term and long-term trends can be simultaneously viewed. In previous studies [14,19,20], only a limited number of countries were included for forecasting. Our system contains 171 countries and provides information about policy measures. Further, data visualization, statistical and deep learning for incidence forecasting, and customized ranking are possible. Based on their objectives, users can select country names and time periods. Similar cultural backgrounds, neighboring geographical characteristics, and high-frequency trading may also serve as attractive features. In particular, a declined ranking is calculated by our system to explore the effectiveness of COVID-19 management strategies implemented in 2020. Thus, the CPAIS is a comprehensive AI-based service that is available on the internet. It relies on big data and offers data visualization, deep learning–based prediction, and customized comparison. This system can be used to investigate COVID-19 progression trends.

Figure 6. Diagram of the long short-term memory neural network with three functional gates.



Limitations

To the best of our knowledge, this is the first web-based machine learning system that can explore variations, trends, and forecasts related to the COVID-19 pandemic across 171 countries. This pilot system still has several limitations. First, this database relies heavily on the source databases and shares similar limitations with the source databases. For example, the source databases did not consider the number of COVID-19 patients that were traveling internationally, and this may result in inaccurate analysis for a small number of countries. However, we think that the number of COVID-19 patients who were traveling internationally is small, as most countries imposed COVID-19–negative tests or proof of vaccination before allowing the traveler into the country. Second, the CPAIS cannot be updated daily if the source databases are not updated. For example, at present, the number of recoveries in the United States was last updated on December 14, 2020. So the number of recoveries in the United States may not be accurate. Finally, since the main purpose of this platform is to consolidate raw data retrieved from various databases and associated measures of pandemic policy implementation, we remind the reader to use text mining, local reports, and information retrieved from the medical system of a given country for further assessment.

Conclusions

In general, the CPAIS collects and summarizes information about the COVID-19 pandemic and offers data visualization

XSL•FO
RenderX

and deep learning–based prediction. It may be a useful and consequential reference resource for any serious outbreak or epidemic that may occur in the future. In addition, information about the vaccine is also stored in our system. It may be used to evaluate the efficacy of the vaccine in different countries in the future. Moreover, the 2-week machine learning forecasts may serve as warning signs and highlight current trends in the epidemic that have been made apparent by AI techniques. To conclude, the CPAIS can be used to summarize several factors that can influence the effectiveness of epidemic prevention and predict the next serious outbreak.

Acknowledgments

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.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hyperparameters, packages, and function code for each model as well as the dynamic heat maps with policy measures for New Zealand, Vietnam, and Taiwan.
[PDF File (Adobe PDF File), 2076 KB - jmir_v23i5e27806_app1.pdf]

References

1. Archived: WHO Timeline - COVID-19. World Health Organization. 2020 Apr 27. URL: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 [accessed 2021-02-01]
2. Origin of SARS-CoV-2. Geneva, Switzerland: World Health Organization; 2020 Mar 26. URL: https://apps.who.int/iris/bitstream/handle/10665/332197/WHO-2019-nCoV-FAQ-Virus_origin-2020.1-eng.pdf [accessed 2021-01-12]
3. Naming the coronavirus disease (COVID-19) and the virus that causes it. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it [accessed 2021-02-16]
4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. 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 2021-01-05]
5. Weekly epidemiological update - 29 December 2020. World Health Organization. 2020 Dec 29. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update---29-december-2020 [accessed 2021-02-14]
6. Imtyaz A, Haleem A, Javaid M. Analysing governmental response to the COVID-19 pandemic. J Oral Biol Craniofac Res 2020;10(4):504-513 [FREE Full text] [doi: 10.1016/j.jobcr.2020.08.005] [Medline: 32834983]
7. Kadi N, Khelfaoui M. Population density, a factor in the spread of COVID-19 in Algeria: Statistic study. Bull Natl Res Cent 2020;44(1):138 [FREE Full text] [doi: 10.1186/s42269-020-00393-x] [Medline: 32843835]
8. 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. J Racial Ethn Health Disparities 2021 Jun;8(3):732-742 [FREE Full text] [doi: 10.1007/s40615-020-00833-4] [Medline: 32875535]
9. Prata DN, Rodrigues W, Bermejo PH. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. Sci Total Environ 2020 Aug 10;729:138862 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138862] [Medline: 32361443]
10. Khalatbari-Soltani S, Cumming RG, Delpierre C, Kelly-Irving M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. J Epidemiol Community Health 2020 Aug;74(8):620-623 [FREE Full text] [doi: 10.1136/jech-2020-214297] [Medline: 32385126]
11. Alwan NA, Burgess RA, Ashworth S, Beale R, Bhadelia N, Bogaert D, et al. Scientific consensus on the COVID-19 pandemic: We need to act now. Lancet 2020 Oct;396(10260):e71-e72. [doi: 10.1016/s0140-6736(20)32153-x]
12. Yu CS, Lin YJ, Lin CH, Lin SY, Wu JL, Chang SS. Development of an online health care assessment for preventive medicine: A machine learning approach. J Med Internet Res 2020 Jun 05;22(6):e18585 [FREE Full text] [doi: 10.2196/18585] [Medline: 32501272]
13. Lin YJ, Lin CH, Wang ST, Lin SY, Chang SS. Noninvasive and convenient screening of metabolic syndrome using the controlled attenuation parameter technology: An evaluation based on self-paid health examination participants. J Clin Med 2019 Oct 24;8(11):1775 [FREE Full text] [doi: 10.3390/jcm8111775] [Medline: 31653028]

XSL•FO
RenderX

14. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Niakan Kalhori SR. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
15. Yeung AYS, Roewer-Despres F, Rosella L, Rudzicz F. Machine learning-based prediction of growth in confirmed COVID-19 infection cases in 114 countries using metrics of nonpharmaceutical interventions and cultural dimensions: Model development and validation. J Med Internet Res 2021 Apr 23;23(4):e26628 [FREE Full text] [doi: 10.2196/26628] [Medline: 33844636]
16. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020 Mar;12(3):165-174 [FREE Full text] [doi: 10.21037/jtd.2020.02.64] [Medline: 32274081]
17. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. Med Image Anal 2021 Jan;67:101860 [FREE Full text] [doi: 10.1016/j.media.2020.101860] [Medline: 33171345]
18. Chen CM, Jyan HW, Chien SC, Jen HH, Hsu CY, Lee PC, et al. Containing COVID-19 among 627,386 persons in contact with the Diamond Princess cruise ship passengers who disembarked in Taiwan: Big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
19. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos Solitons Fractals 2020 Nov;140:110212 [FREE Full text] [doi: 10.1016/j.chaos.2020.110212] [Medline: 32839642]
20. Shastri S, Singh K, Kumar S, Kour P, Mansotra V. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. Chaos Solitons Fractals 2020 Nov;140:110227 [FREE Full text] [doi: 10.1016/j.chaos.2020.110227] [Medline: 32843824]
21. Kostoff RN, Briggs MB, Porter AL, Spandidos DA, Tsatsakis A. [Comment] COVID-19 vaccine safety. Int J Mol Med 2020 Nov;46(5):1599-1602 [FREE Full text] [doi: 10.3892/ijmm.2020.4733] [Medline: 33000193]
22. Phillips N. The coronavirus is here to stay - Here's what that means. Nature 2021 Feb;590(7846):382-384 [FREE Full text] [doi: 10.1038/d41586-021-00396-2] [Medline: 33594289]
23. Scudellari M. How the pandemic might play out in 2021 and beyond. Nature 2020 Aug;584(7819):22-25 [FREE Full text] [doi: 10.1038/d41586-020-02278-5] [Medline: 32760050]
24. Baker MG, Wilson N, Anglemyer A. Successful elimination of Covid-19 transmission in New Zealand. N Engl J Med 2020 Aug 20;383(8):e56. [doi: 10.1056/nejmc2025203]
25. Kim S, Castro MC. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). Int J Infect Dis 2020 Sep;98:328-333 [FREE Full text] [doi: 10.1016/j.ijid.2020.07.004] [Medline: 32634584]
26. CPAIS: COVID-19 Pandemic AI System. URL: https://covid19.mldoctor.com.tw [accessed 2021-05-03]
27. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]
28. 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]
29. Johns Hopkins University Center for Systems Science and Engineering. CSSEGISandData/COVID-19: Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. GitHub. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-05-01]
30. Blavatnik School of Government. OxCGRT/covid-policy-tracker: Systematic dataset of Covid-19 policy, from Oxford University. GitHub. URL: https://github.com/OxCGRT/covid-policy-tracker [accessed 2020-06-15]
31. World Population Prospects 2019, Online Edition. Rev 1. New York, NY: United Nations, Department of Economic and Social Affairs, Population Division; 2019. URL: https://population.un.org/wpp/Download/Standard/Population/ [accessed 2021-04-06]
32. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast 2006 Oct;22(4):679-688. [doi: 10.1016/j.ijforecast.2006.03.001]
33. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 2nd edition. Melbourne, Australia: OTexts; 2018.
34. Hamilton JD. Time Series Analysis. Princeton, NJ: Princeton University Press; 1994.
35. Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. New York, NY: Springer; 2002.
36. Hyndman RJ, Khandakar Y. Automatic time series forecasting: The forecast package for R. J Stat Softw 2008 Jul;27(3):1-22 [FREE Full text] [doi: 10.18637/jss.v027.i03]
37. Wang X, Smith K, Hyndman R. Characteristic-based clustering for time series data. Data Min Knowl Discov 2006 May 16;13(3):335-364. [doi: 10.1007/s10618-005-0039-x]
38. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
39. Zell A. Simulation Neuronaler Netze. Boston, MA: Addison-Wesley Bonn; 1994.
40. Ripley BD. Pattern Recognition and Neural Networks. Cambridge, UK: Cambridge University Press; 1996.
41. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th edition. New York, NY: Springer; 2002.

XSL•FO
RenderX

42. Trevor H, Robert T, Jerome F. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2009.
43. Rosenblatt F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington, DC: Spartan Books; 1961.
44. Haykin S. Neural Networks: A Comprehensive Foundation. 2nd edition. Upper Saddle River, NJ: Prentice Hall; 1998.
45. Ord K, Fildes R, Kourentzes N. Principles of Business Forecasting. 2nd edition. New York, NY: Wessex Press Publishing Co; 2017.
46. Kourentzes N, Barrow DK, Crone SF. Neural network ensemble operators for time series forecasting. Expert Syst Appl 2014 Jul 12;41(9):4235-4244. [doi: 10.1016/j.eswa.2013.12.011]
47. Crone S, Kourentzes N. Feature selection for time series prediction – A combined filter and wrapper approach for neural networks. Neurocomputing 2010 Jun;73(10-12):1923-1936. [doi: 10.1016/j.neucom.2010.01.017]
48. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
49. Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Germany: Springer-Verlag; 2012.
50. Gal Y, Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks. In: Proceedings of the 30th Conference on Neural Information Processing Systems. 2016 Presented at: 30th Conference on Neural Information Processing Systems; December 5-10, 2016; Barcelona, Spain URL: https://papers.nips.cc/paper/2016/file/076a0c97d09cf1a0ec3e19c7f2529f2b-Paper.pdf
51. Ramasubramanian K, Singh A. Deep learning using Keras and TensorFlow. In: Machine Learning Using R: With Time Series and Industry-Based Use Cases in R. New York, NY: Apress; 2019:667-688.
52. TensorFlow. URL: https://www.tensorflow.org/ [accessed 2021-01-04]
53. Covid Performance Index. The Lowy Institute. URL: https://interactives.lowyinstitute.org/features/covid-performance/ [accessed 2021-02-15]
54. Papoulis A, Pillai SU. Probability, Random Variables and Stochastic Processes. 4th edition. New York, NY: McGraw-Hill; 2002.
55. Triacca U. The Wold decomposition theorem. International Doctoral Program in Economics, Scuola Superiore Sant'Anna. URL: http://www.phdeconomics.sssup.it/documents/Lesson11.pdf [accessed 2021-02-19]

Abbreviations

AI: artificial intelligence
ARIMA: autoregressive integrated moving average
CPAIS: COVID-19 Pandemic AI System
CSSE: Center for Systems Science and Engineering
FNN: feedforward neural network
LSTM: long short-term memory
MAE: mean absolute error
MAPE: mean absolute percentage error
ME: mean error
MLP: multilayer perceptron
MPE: mean percentage error
NNAR: neural network autoregression model
OxCGRT: Oxford COVID-19 Government Response Tracker
RMSE: root mean square error
RNN: recurrent neural network
Sass: Syntactically Awesome Style Sheets
SQL: Structured Query Language
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.03.21; peer-reviewed by MT Lee, YL Chan, JA Benítez-Andrades; comments to author 04.04.21; revised version received 12.04.21; accepted 23.04.21; published 20.05.21.

Please cite as:
Yu CS, Chang SS, Chang TH, Wu JL, Lin YJ, Chien HF, Chen RJ
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
URL: https://www.jmir.org/2021/5/e27806
doi:10.2196/27806
PMID:33900932

©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), 20.05.2021. 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

Characterization of an Open-Access Medical News Platform's Readership During the COVID-19 Pandemic: Retrospective Observational Study

Alex K Chan[1,2], BSc; Constance Wu[3], BA; Andrew Cheung[1,4], MD; Marc D Succi[2,3,5], MD

[1]Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada

[2]Medically Engineered Solutions in Healthcare Incubator, Innovation in Operations Research Center, Massachusetts General Hospital, Boston, MA, United States

[3]Harvard Medical School, Boston, MA, United States

[4]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[5]Department of Radiology, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Marc D Succi, MD
Department of Radiology
Massachusetts General Hospital
55 Fruit Street
Boston, MA, 02114
United States
Phone: 1 6179359144
Email: msucci@mgh.harvard.edu

Abstract

Background: There are many alternatives to direct journal access, such as podcasts, blogs, and news sites, that allow physicians and the general public to stay up to date with medical literature. However, there is a scarcity of literature that investigates the readership characteristics of open-access medical news sites and how these characteristics may have shifted during the COVID-19 pandemic.

Objective: This study aimed to assess readership and survey data to characterize open-access medical news readership trends related to the COVID-19 pandemic and overall readership trends regarding pandemic-related information delivery.

Methods: Anonymous, aggregate readership data were obtained from 2 Minute Medicine, an open-access, physician-run medical news organization that has published over 8000 original, physician-written texts and visual summaries of new medical research since 2013. In this retrospective observational study, the average number of article views, number of actions (defined as the sum of the number of views, shares, and outbound link clicks), read times, and bounce rates (probability of leaving a page in <30 s) were compared between COVID-19 articles published from January 1 to May 31, 2020 (n=40) and non–COVID-19 articles (n=145) published in the same time period. A voluntary survey was also sent to subscribed 2 Minute Medicine readers to further characterize readership demographics and preferences, which were scored on a Likert scale.

Results: COVID-19 articles had a significantly higher median number of views than non–COVID-19 articles (296 vs 110; U=748.5; P<.001). There were no significant differences in average read times (P=.12) or bounce rates (P=.12). Non–COVID-19 articles had a higher median number of actions than COVID-19 articles (2.9 vs 2.5; U=2070.5; P=.02). On a Likert scale of 1 (strongly disagree) to 5 (strongly agree), our survey data revealed that 65.5% (78/119) of readers agreed or strongly agreed that they preferred staying up to date with emerging literature about COVID-19 by using sources such as 2 Minute Medicine instead of journals. A greater proportion of survey respondents also indicated that open-access news sources were one of their primary sources for staying informed (86/120, 71.7%) compared to the proportion who preferred direct journal article access (61/120, 50.8%). The proportion of readers indicating they were reading one or less full-length medical studies a month were lower following introduction to 2 Minute Medicine compared to prior (21/120, 17.5% vs 38/120, 31.6%; P=.005).

Conclusions: The readership significantly increased for one open-access medical literature platform during the pandemic. This reinforces the idea that open-access, physician-written sources of medical news represent an important alternative to direct journal access for readers who want to stay up to date with medical literature.

XSL•FO
RenderX

(J Med Internet Res 2021;23(5):e26666) doi:10.2196/26666

KEYWORDS

COVID-19; internet; medical news; text summaries; readership trends; news; media; open access; literature; web-based health information; survey; cross-sectional; trend

Introduction

On March 11, 2020, COVID-19 was declared a pandemic by the World Health Organization roughly 11 weeks after the report of the first detected COVID-19 case [1]. With its high transmissibility and global impact, COVID-19 has attracted tremendous interest from researchers, clinicians, and the general population worldwide. Within the first 3 months of the disease's discovery, bibliometric analyses indicated that there was already a greater number of peer-reviewed COVID-19 studies published than the combined number of articles on SARS (severe acute respiratory syndrome) and MERS (Middle East respiratory syndrome) published during these diseases' first year of discovery [2,3]. This has also resulted in significant interest from the mass media, as a substantial increase in COVID-19–related news was observed in the early months of 2020 [4]. Prior literature has found that overall news consumption increases significantly during a national crisis, and this appears to be the case for the COVID-19 outbreak, as there is evidence suggesting that information-seeking behavior increased following the declaration of the pandemic [5-7].

Compared to the past, many alternatives to direct journal access and mainstream media now exist for clinicians and the general public who want to stay informed about medical news and research, including, but not limited to, social media, blogs, journal newsletters, open-access medical news sites, and podcasts. Although there is evidence that suggests that differences in the use of these alternative sources can have subsequent downstream effects on consumers' health behaviors, there is little data on the characterization of the consumption of these sources both prior to and following the start of the pandemic [8]. Specifically, there is a notable scarcity of investigations examining the readership trends and characteristics of open-access medical news organizations.

To further characterize the consumption of medical literature during the COVID-19 pandemic, this study sought to investigate readership trends for one physician-run, open-access medical news organization and readership preferences regarding the delivery of COVID-19–related research and information.

Methods

Readership Data

Aggregate, anonymous data for this retrospective observational study was obtained from 2 Minute Medicine Inc [9]. 2 Minute Medicine is a free, open-access medical news organization that publishes daily, physician-written texts and visual summaries of new medical research. It has published over 8000 summaries since 2013. Articles published on the website from January 1 to May 31, 2020, were included in our analysis. In that time span, 40 articles were published about COVID-19–related research, and 145 articles were published about non–COVID-19 medical research. Overall daily website traffic, average view times, the average number of actions, and average bounce rates were longitudinally characterized for the study period.

Web-Based Survey Data

A web-based survey was sent to 4221 readers who opted into the website's free, daily electronic mailing list. Nonidentifying demographic data, including age, sex, respondents' level of education, and respondents' field of work or study, were gathered. Survey items were created with the intention of directly addressing identified gaps in literature regarding respondents' perspectives and the relative use of alternative sources, as secondary outcome validation was beyond the scope of this study. Descriptive data were gathered on readers' behaviors and preferences via Likert scales that ranged from 1 (strongly disagree) to 5 (strongly agree). The key statements used in the study included items such as "I prefer using open access medical news sites such as www.2minutemedicine.com to stay up to date with research related to COVID-19." Users were also asked to indicate all of the primary sources that they were using to stay informed about COVID-19–related literature, including, but not limited to, direct journal access, journal newsletters, mainstream media, social media, and open-access medical news sites such as 2 Minute Medicine. Other readership trends were also briefly investigated with the survey, as users indicated their frequency of reading full-length original journal articles both prior to and following their introduction to 2 Minute Medicine.

Ethics

This study was compliant with the Health Insurance Portability and Accountability Act and exempt from review by the institutional review board.

Statistical Analysis

All statistical analyses were performed using SPSS, version 26 (IBM Corporation). The visual inspection of a histogram, normal quantile-quantile plot, and box plot and the Kolmogorov-Smirnov test were completed to determine if the data followed a normal distribution [10]. If the data were not normally distributed, the Mann-Whitney *U* test was used to compare the average number of views (over the first 2 weeks following publication on the website), number of actions (defined as the sum of the number of views, shares, and outbound link clicks), read times (in seconds), and bounce rates (defined as the probability of readers leaving the webpage in <30 s) for each article [11]. In this study, actions, read times, and bounce rates were used as measures of reader engagement.

Results

A total of 121 responses were obtained over a 7-week period (response rate: 121/4421, 2.7%). Demographic characteristics of respondents are summarized in Table 1. The majority of

XSL•FO
RenderX

respondents reported that they worked in the health care industry (114/121, 94.2%). These respondents included undergraduate students, graduate students, medical students, medical residents, and licensed physicians, and they made up the majority of the respondents.

Table 1. Demographic characteristics of respondents.

Characteristics	Value, n (%)
Sex	
Male	66 (55)
Female	54 (45)
Age group (years)	
<20	1 (0.8)
20-29	43 (35.8)
30-39	37 (30.8)
40-49	6 (5)
50-59	7 (5.8)
60-69	10 (8.3)
>69	9 (7.5)
Did not indicate	6 (5)
Education	
Less than or equal to high school graduate	3 (2.5)
Some college or university education	13 (10.8)
Undergraduate degree	18 (15)
Master's degree	24 (20)
Doctoral or professional degree	62 (51.6)
Field of work or study	
Health care worker	116 (95.9)
Other	5 (4.1)
Licensed physician	26 (21.4)
Licensed allied health professional	13 (10.7)
Licensed pharmacist	7 (5.7)
Medical or surgical fellow	6 (4.9)
Medical or surgical resident	34 (28.1)
Pharmacy resident	1 (0.8)
Medical student	6 (4.9)
Graduate student	10 (8.2)
Undergraduate student	9 (7.3)
Pharmacy student	2 (1.7)
Unspecified or other	6 (4.9)

Over the study period, COVID-19 articles had a total of 68,129 views (mean 1792, SD 6491), while articles not related to the pandemic accrued 19,650 views (mean 137, SD 122). Our analysis via the Mann-Whitney *U* test revealed that COVID-19 articles published within the observed time frame had a significantly higher median number of views than non–COVID-19 articles (296 vs 110; U=748.5; P<.001). There was no difference in average view times or bounce rates between the two types of articles. Non–COVID-19 articles had a significantly higher median number of actions than COVID-19 articles (2.9 vs 2.5; U=2070.5; P=.02).

The number of daily visitors over the study period are displayed in Figure 1. The average mean daily visitor count was 1724 (SD 1549). Daily visitor counts were relatively stable from January 1 to February 29, 2020 (mean 1298, SD 292), but a significant peak in daily readership was observed in the month of March (mean 2816, SD 3134), and a peak visitor count of 12,806 daily

XSL•FO
RenderX

readers was observed on March 24. The mean daily visitor count from April 1 to May 31 was 1586 (SD 437). The average visit time per page over the study period was 168 s (SD 16 s), and this is shown in Figure 2. The lowest average daily read times were found on March 23 (116 s), March 24 (115 s), and March 25 (120 s).

Figure 1. Daily visitor traffic from January 1 to May 30, 2020 compared to the baseline average.

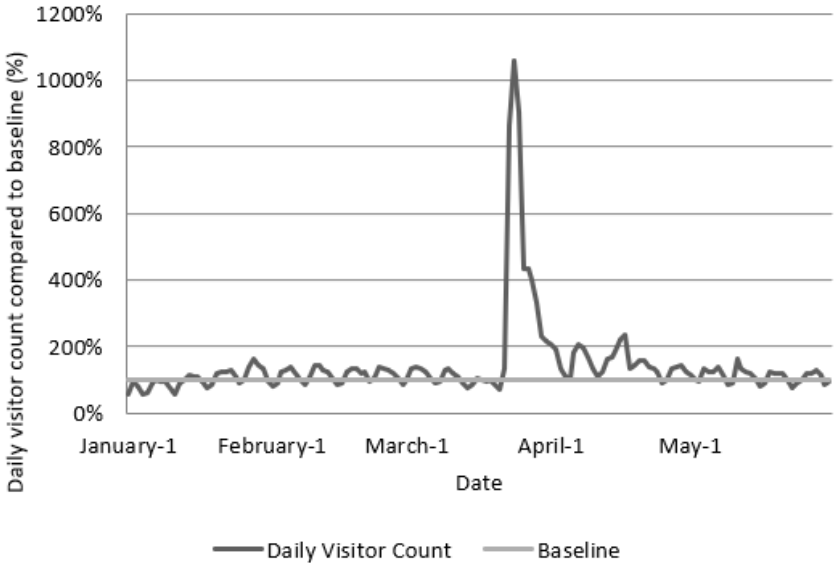
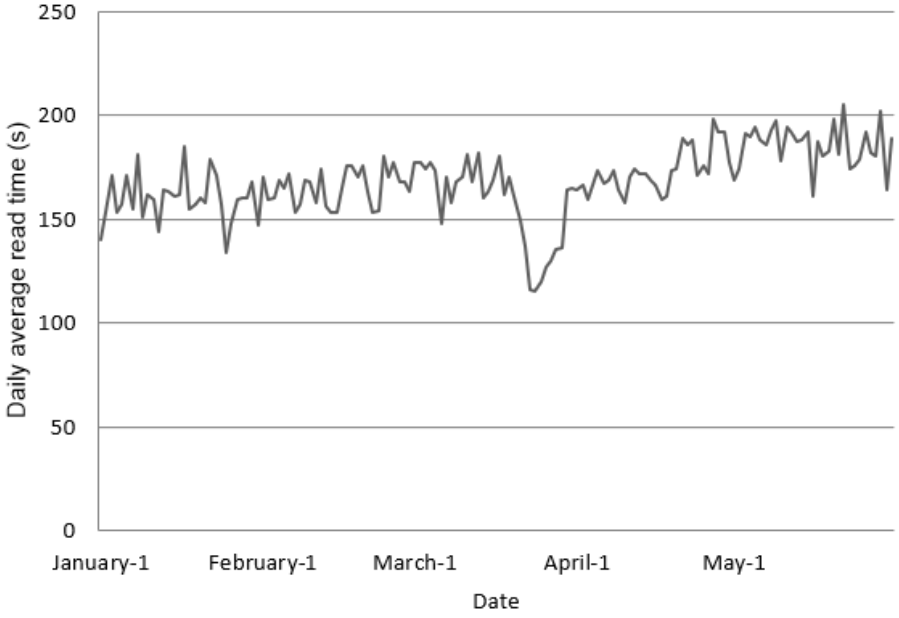


Figure 2. Daily average read times for articles from January 1 to May 30, 2020.



In terms of survey responses, on a Likert scale of 1 (strongly disagree) to 5 (strongly agree), data from 119 responses (2 respondents did not answer) revealed that 65.5% (78/119) of readers agreed or strongly agreed that they preferred staying up to date with emerging literature about COVID-19 by using sources such as 2 Minute Medicine instead of journals, as shown in Table 2.

When respondents were asked to indicate all of the sources they used to stay up to date with COVID-19 literature, a greater proportion of survey takers indicated that open-access news sources were one of their primary means of staying informed (86/120, 71.7%) compared to the proportion who preferred direct journal article access (61/120, 50.8%), as shown in Table 3. Journal newsletter updates (70/120, 58.3%), social media sources (53/120, 44.2%), and mainstream news media sources (46/120, 38.3%) were other recognized sources of information. The following five other sources were also recognized as sources of information: work or school announcements, podcasts, institutional newsletters, and search engine notifications. In terms of the impact of 2 Minute Medicine on information consumption, a Fisher exact test demonstrated that the proportion of readers who indicated that they were reading 1 or fewer full-length medical studies per month decreased after their introduction to 2 Minute Medicine (21/120, 17.5% vs 38/120, 31.6%; *P*=.005).

XSL•FO
RenderX

Table 2. Likert scale data for the question "I prefer staying up to date with emerging research surrounding COVID-19 by using sources such as 2 Minute Medicine versus reading the articles themselves."

Selection	Responses, n (%)
Strongly Disagree	3 (2.5)
Disagree	10 (8.4)
Neutral	28 (23.5)
Agree	39 (32.8)
Strongly Agree	39 (32.8)

Table 3. Sources identified by readers as their primary means for staying informed with COVID-19–related research.

Source	Responses, n (%)
Open-access medical news sites	86 (71.7)
Journal newsletters	70 (58.3)
Direct journal access	61 (50.8)
Social media	53 (44.2)
Mainstream media	46 (38.3)
Research newsletters	20 (16.7)
Research aggregator services	16 (13.3)
Other sources	5 (4.2)

Discussion

Principal Results

The results in this study show that average view counts for COVID-19–related articles were significantly higher than those of articles that covered other medical news during the pandemic (P<.001). Although this suggests people's overall interest in COVID-19–related news was greater than their interest in other medical news, there were no associated findings for measures of audience engagement, including average view times or bounce rates. Interestingly however, the average number of actions, which consisted of the number of views, shares, and outbound link clicks per article, were found to be significantly higher for non–COVID-19 medical news articles (P=.02), which had an average of nearly 3 actions. We hypothesized that this was due to these articles tending to cover subspecialties or other niche areas of medicine, thereby attracting readers that may be more heavily involved in these fields and have a higher probability of reading primary source articles.

Longitudinal trends of daily visitor counts indicated relatively stable overall site traffic for the first 3 months of 2020. A large spike in daily readership was observed between March 23 to March 26. This was followed by variable daily site traffic before it returned to relative stability in the third and fourth weeks of April. Interestingly, while prior literature has documented increased information-seeking behavior in the first 1 or 2 days following local announcements of outbreaks, this large spike in site traffic appeared to be primarily attributable to a site that published summaries covering research on hydroxychloroquine and its potential therapeutic application for treating COVID-19 [7]. Daily visitor counts dropped in the following weeks but remained above the baseline average. This is possibly indicative of readership retention. Furthermore, while average read times remained fairly stable with an SD of less than 10% over the study period, a substantial decrease in average read times coincided with this peak in visitor count; the decreased average read time was 30.3%. Although this was not confirmed in our site analysis, we believed that this decrease in average read time may be reflective of the site's larger reach to the general public compared to the normal demographic of the site's readers—health care professionals. Overall, our results indicate that secondary sources such as medical news sites may have readership data that are more heavily influenced by the gravity and popularity of individual research articles than by other potential factors, such as infection count or other trends related to the pandemic. However, this will need to be further validated in future studies. To our knowledge, these findings represent the first characterization of open-access medical news readership trends during an international health crisis.

In our review of the survey results, the survey data collected appeared to corroborate the findings of the site data comparisons. Unsurprisingly, the majority of respondents (114/121, 94.2%) to the survey, who were subscribers of 2 Minute Medicine, were primarily composed of those directly involved in health care. The survey results indicated that users preferred open-access medical news sources such as 2 Minute Medicine as their primary source of information about pandemic-related research. This is based on both the overall agreement responses on the Likert scale and the fact that a greater proportion of respondents indicated that such sources were their primary news source compared to the proportion who used journals, journal newsletters, and mainstream media as their primary news source. The transition to using sources such as open-access new sources and newsletters may be increasing in prevalence due to the explosive growth in the volume of

XSL•FO
RenderX

scientific literature being published per year [11]. The use of these alternative sources may provide readers with a means to sort through such literature and identify research with the highest impact to guide their reading. Interestingly, the use of 2 Minute Medicine appeared to influence the respondents' overall consumption of medical literature, as readers indicated that they read a greater number of articles via direct journal access following their introduction to the website. In terms of the demographic data acquired in the study, it is of note that despite people's increased interest in 2 Minute Medicine and its articles covering COVID-19–related research, the vast majority of subscribers to the website continue to be composed of respondents who are directly involved in the health care field. As such, further investigations may be needed to characterize the readership trends and preferences of the general public.

To our knowledge, this is the first study that describes open-access medical news readership trends and preferences for sources alternative to journals during the COVID-19 pandemic. During an outbreak in which an abundance of research is being published while there is high global interest in the pandemic, having a better understanding of how information is consumed by both health care workers and the general public may provide a better understanding of the behaviors associated with the pandemic. Prior literature has demonstrated that different sources of information are at variable risk of misinformation, which may have subsequent impacts on consumer behavior [12-16]. For example, a prior study conducted by Allington et al [8] found a positive correlation between the use of social media and the frequency of COVID-19–related conspiracy theories and a negative relationship between COVID-19–related health-protective behaviors and the use of social media.

Limitations

In terms of the limitations of this study, as the survey was only sent to subscribers of 2 Minute Medicine—the vast majority of whom are in the field of health care—the generalizability of our findings to the general public are unknown. The sample used for the survey also invariably skewed data to favor preferences toward open-access medical sites. This was a result of sampling bias, as respondents were active subscribers of the organization. Further studies should more broadly survey health care professionals to repeat our findings. It is also recognized that the survey used in this study lacks any validity evidence, and the wording of questions may have been framed positively toward 2 Minute Medicine, which may have introduced bias in the collection of results. Additionally, although the demographics of survey respondents were gathered, individual demographic data on site visitors were not obtained for the purposes of this study, which led to a lack of understanding of the visitors to the site itself. Finally, as survey data on the frequency of reading full-length original studies were not gathered prospectively, this study can only provide limited information regarding the interaction between open-access medical news consumption and original research consumption.

Conclusions

The results in this study indicate that readership on an open-access medical news site significantly increased as a result of the COVID-19 pandemic. Overall site traffic remained relatively stable in the first 2 months of the study period, and a substantial spike in traffic occurred in mid-March 2020. Based on the comparisons between article view counts and survey data, it appears as though open-access medical news sites may represent an important source for physicians and other health care workers who want to stay up to date with relevant COVID-19 literature. The introduction to such sites may also have subsequent impacts on overall medical literature consumption and increase the frequency of direct journal access.

Conflicts of Interest

All authors declare a financial interest (whether in compensation, equity, or both) in 2 Minute Medicine Inc, which provided the publishing platform, aggregate readership data, and readership audience for testing the hypothesis and surveying the readers.

References

1. Archived: WHO timeline - COVID-19. World Health Organization. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-05-06]
2. Chiu W, Huang J, Ho Y. Bibliometric analysis of Severe Acute Respiratory Syndrome-related research in the beginning stage. Scientometrics 2004;61(1):69-77 [FREE Full text] [doi: 10.1023/B:SCIE.0000037363.49623.28] [Medline: 32214553]
3. Zyoud SH. Global research trends of Middle East respiratory syndrome coronavirus: a bibliometric analysis. BMC Infect Dis 2016 Jun 07;16:255 [FREE Full text] [doi: 10.1186/s12879-016-1600-5] [Medline: 27267256]
4. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
5. Dotan J, Cohen AA. Mass media use in the family during war and peace. Communic Res 2016 Jun 29;3(4):393-402. [doi: 10.1177/009365027600300403]
6. Althaus SL. American news consumption during times of national crisis. PS Polit Sci Polit 2002 Sep 11;35(3):517-521. [doi: 10.1017/s104909650200077x]

XSL•FO
RenderX

7. 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]
8. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7 [FREE Full text] [doi: 10.1017/S003329172000224X] [Medline: 32513320]
9. Concise. Curated. Medical News. 2 Minute Medicine. URL: https://www.2minutemedicine.com/ [accessed 2020-12-12]
10. Lilliefors HW. On the Kolmogorov-Smirnov Test for normality with mean and variance Unknown. J Am Stat Assoc 1967 Jun;62(318):399-402. [doi: 10.1080/01621459.1967.10482916]
11. Govindarajulu Z. Robustness of Mann-Whitney-Wilcoxon test for scale to dependence in the variables. In: Block HW, Sampson AR, Savits TH, editors. Institute of Mathematical Statistics Lecture Notes: Monograph Series, Volume 16, Topics in Statistical Dependence. Hayward, California: Institute of Mathematical Statistics; 1990:237-250.
12. Landhuis E. Scientific literature: Information overload. Nature 2016 Jul 21;535(7612):457-458. [doi: 10.1038/nj7612-457a] [Medline: 27453968]
13. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2020-12-12]
14. 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]
15. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
16. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Sep;112(6):320-328 [FREE Full text] [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]

Abbreviations

MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 20.12.20; peer-reviewed by B Barman, P Suber; comments to author 06.01.21; revised version received 11.01.21; accepted 16.04.21; published 25.05.21.

Please cite as:
Chan AK, Wu C, Cheung A, Succi MD
Characterization of an Open-Access Medical News Platform's Readership During the COVID-19 Pandemic: Retrospective Observational Study
J Med Internet Res 2021;23(5):e26666
URL: https://www.jmir.org/2021/5/e26666
doi:10.2196/26666
PMID:33866307

©Alex K Chan, Constance Wu, Andrew Cheung, Marc D Succi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2021. 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 Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups Among Patients with COVID-19: Observational Longitudinal Study

Julián Benito-León[1*], MD, PhD; Mª Dolores del Castillo[2*], PhD; Alberto Estirado[3], BSc; Ritwik Ghosh[4], MD; Souvik Dubey[5], MD, DM; J Ignacio Serrano[2*], PhD

[1]Department of Neurology, University Hospital "12 de Octubre", Madrid, Spain

[2]Neural and Cognitive Engineering Group, Center for Automation and Robotics, CSIC-UPM, Arganda del Rey, Spain

[3]HM Hospitales, Madrid, Spain

[4]Department of General Medicine, Burdwan Medical College and Hospital, Burdwan, India

[5]Department of Neuromedicine, Bangur Institute of Neurosciences, Kolkata, India

[*]these authors contributed equally

Corresponding Author:
Julián Benito-León, MD, PhD
Department of Neurology
University Hospital "12 de Octubre"
Avenida de Córdoba s/n
Madrid, 28041
Spain
Phone: 34 639154069
Email: jbenitol67@gmail.com

Abstract

Background: Early detection and intervention are the key factors for improving outcomes in patients with COVID-19.

Objective: The objective of this observational longitudinal study was to identify nonoverlapping severity subgroups (ie, clusters) among patients with COVID-19, based exclusively on clinical data and standard laboratory tests obtained during patient assessment in the emergency department.

Methods: We applied unsupervised machine learning to a data set of 853 patients with COVID-19 from the HM group of hospitals (HM Hospitales) in Madrid, Spain. Age and sex were not considered while building the clusters, as these variables could introduce biases in machine learning algorithms and raise ethical implications or enable discrimination in triage protocols.

Results: From 850 clinical and laboratory variables, four tests—the serum levels of aspartate transaminase (AST), lactate dehydrogenase (LDH), C-reactive protein (CRP), and the number of neutrophils—were enough to segregate the entire patient pool into three separate clusters. Further, the percentage of monocytes and lymphocytes and the levels of alanine transaminase (ALT) distinguished cluster 3 patients from the other two clusters. The highest proportion of deceased patients; the highest levels of AST, ALT, LDH, and CRP; the highest number of neutrophils; and the lowest percentages of monocytes and lymphocytes characterized cluster 1. Cluster 2 included a lower proportion of deceased patients and intermediate levels of the previous laboratory tests. The lowest proportion of deceased patients; the lowest levels of AST, ALT, LDH, and CRP; the lowest number of neutrophils; and the highest percentages of monocytes and lymphocytes characterized cluster 3.

Conclusions: A few standard laboratory tests, deemed available in all emergency departments, have shown good discriminative power for the characterization of severity subgroups among patients with COVID-19.

(J Med Internet Res 2021;23(5):e25988) doi:10.2196/25988

KEYWORDS

COVID-19; machine learning; outcome; severity; subgroup; emergency; detection; intervention; testing; data set; characterization

XSL•FO
RenderX

Introduction

The COVID-19 pandemic has brought to light the scarcity of health care resources worldwide [1]. One of the main challenges faced by health care systems while tackling this pandemic is the lack of affordable, accurate, and simple information that can allow clinicians to predict the evolution of the patients' disease sooner, upon admission to the hospital. This information might help clinicians to make early decisions regarding arrangement and organization of medical resources, as well as early interventions to improve the health outcomes of these patients.

The exhaustive and inefficiently structured amount of health data available does not permit parametric modeling in an easy way. To overcome this issue, machine learning techniques have recently been identified as promising tools in data analysis for individual class prediction, allowing us to deal with a great number of variables simultaneously and observe inherent disease-related patterns in the data [2].

Machine learning for health care is a key discipline aimed to translate large health data sets into operative knowledge in different medical fields [3-7]. The methods of this artificial intelligence paradigm can be classified as supervised or unsupervised, based on the underlying strategy used [8]. In inductive or supervised machine learning, the method builds a general class description of the target categories from a set of previously categorized examples [8]. In general, supervised learning methods are used to design classifiers from labeled samples that predict the class of an unseen new sample [8]. In the field of medicine, these methods have been applied to find prognostic and predictive biomarkers [9]. On the other hand, in unsupervised machine learning, the goal is to find the class or classes that cover the sample [8]. These methods permit the discovery of the underlying structure and relationships among unlabeled samples [8]. Unsupervised clustering techniques can obtain groups of samples so that the intrasimilarity within each group is maximized, while intersimilarity between groups is minimized [8]. They are usually applied in medicine to identify homogeneous groups of patients based on their medical records and relationships between clinical manifestations and therapeutic responses, or to detect sets of coexpressed genes, among other applications [10,11].

There are several research reports using COVID-19 data sets, which focus on predicting the patients' mortality or severity by mainly using regression modeling from labeled clinical records [12-17]. Further, in a multicenter study, using supervised machine learning, a personalized COVID-19 mortality risk score for hospitalized patients upon admission has been proposed [18]; however, in that study [18], the reason for choosing only a subset of the recorded clinical variables to build their model was not explained. Therefore, the algorithm might have been biased, even by the expert's knowledge. In all of these studies [12-18] and in a study based on cluster analysis [19], demographics, such as age and sex, were considered as key variables in their prediction models. By contrast, these variables were deliberately excluded from the training data set in this study, in which we used an unsupervised machine learning method for data handling.

Health agencies recommend that clinical decisions should be made based on an individual's biological age rather than chronological age [20,21]. There are multiple physiological and molecular markers for estimation of biological aging that can predict life span [22]. Besides these markers, the heterogeneity of eating habits, physical and mental conditions, and therapeutics have an influence on the overall health state, making biological aging a heterogeneous process too.

Frailty and multi-morbidities, as measures of biological aging, have been found to be risk factors for mortality independent of chronological age in patients with COVID-19 [23]. New procedures for the therapeutic management of COVID-19 are required regardless chronological age [24].

Furthermore, reports about case-fatality rates for COVID-19 categorized by age groups could sentence elderly people not only to social exclusion but also to health care indifference. Considering the elderly population as a highly vulnerable group is a simple and negative stereotype that may even influence decision making in clinical resource management [25].

The prevalence and severity of COVID-19 also varies based on sex, whereby men experience higher mortality than women [26]. The severity of the disease implies that the person may need hospitalization, intensive care support, and mechanical ventilation. However, the medical treatments scheduled during hospitalization or a stay in intensive care are the same for every patient with a severe case of COVID-19, regardless age or sex [27].

Since chronological age as well as sex cannot be considered as pivotal aspects to determine an individual's health status and resilience [28], these should not be key determinants for health care or resource allocation among people suffering from COVID-19. Therefore, predictive models based on intelligent data processing that take into account a patient's age as a major determinant in health care access may be inappropriate and unethical [25].

Demographic variables (ie, age and sex) were not used in the previously published studies for building models on effective treatments based upon sex or age groups or for understanding sex or age differences [12-19]. These predictive models of severity and mortality risk for COVID-19 could be discriminating [29]. For example, consideration of the age of people in the emergency department might discriminate against older people (ie, ageism) regarding access to care, since the decision would be based purely upon the age of the patients rather than their health care needs [30].

The objective of this observational longitudinal study was to identify nonoverlapping severity subgroups (ie, clusters) among patients with COVID-19, using exclusive laboratory tests and clinical data obtained during the first medical contact in the emergency department, by means of unsupervised machine learning techniques. Age and sex were not taken into account to build the subgroups due to the ethical implications. For this purpose, we used the data set collected by the HM group of hospitals (HM Hospitales) in Madrid, Spain [31].

XSL•FO
RenderX

Methods

Data Set

This study is a longitudinal analysis of the data set collected by the HM group of hospitals in Madrid, Spain, in the context of the project Covid Data Save Lives [31]. The information from this data set comes from the electronic health records data system of the seven HM hospitals, located in the Community of Madrid in Spain [31]. This data set contains the anonymized records of 2310 patients, admitted to any of the seven HM hospitals, with a diagnosis of COVID-19 from March 1 to April 24, 2020. The data set includes different interactions within COVID-19 treatment processes, including detailed information on diagnoses, treatments, intensive care unit (ICU) admission, and discharge or death, among many other variables. The data set also includes diagnostic imaging and laboratory tests or records of previous medical care, if any. It also includes the drugs administered to each patient during admission (more than 60,000 records) with the dates corresponding to the first and last administration of each drug, which was identified by its brand name and classification in the Anatomical Therapeutic Chemical codes (ATC5/ATC7). Moreover, laboratory data are also included (398,884 records). Finally, the data set contains the records of the diagnostic and procedural information—coded according to the ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) classification in its latest distributed version—for the patients referred, both for episodes of hospital admission (more than 1600) and for the emergencies (more than 1900) prior to those episodes, if any.

Data Preprocessing

We collected the information for each patient identifier and compiled it into one record. This included age, sex, vital signs in the emergency department, and the need or lack of need of the ICU. COVID-19 symptoms, ICD-10 codes of previous and current conditions, as well as different laboratory tests performed in the emergency department were also recorded. We also calculated, for each patient, the duration in days of the hospital stay, including ICU admission and the days from hospitalization to ICU admission. We also considered the first laboratory tests obtained from the emergency department and grouped all of the ICD-10 codes under the first three characters (ie, first letter and two subsequent numbers) of the code to reduce the number of variables and provide generalization. We codified each ICD-10 feature for inclusion in one of the following groups: *present in emergency department admission*, *not present in emergency department admission*, or *developed during hospital stay*.

Only patients with a discharge reason of *death* or *recovered* were included in the analyses. The patients with a discharge reason of *transferred to another hospital* or *transferred back to the nursing home* (about 3.6% of the total data set) were excluded, since no additional information was available after they left the hospital. We only selected the records (ie, patients) with no missing values on clinical data and laboratory tests, which left a final sample of 853 (37.2% women) patients to be included in our analyses. The mean age of the sample was 67.2 (SD 15.7) years (range 21-106). Each patient had 850 variables in his or her record, including eight variables about demographics, hospitalization stay, and outcome measures; one variable about COVID-19 symptoms; 10 variables about vital signs (eg, temperature, heart rate, oxygen saturation, and systolic and diastolic blood pressure) in the emergency department; 29 laboratory tests from the emergency department (Table 1); 168 ICD-10 codes from the emergency department; and 634 ICD-10 codes during their hospital stay.

The final sample of 853 patients was similar to the excluded sample (n=1457) in terms of age (mean 67.2, SD 15.7 years, vs mean 67.1, SD 17.0 years; $F_{1,2308}$=1.508; P=.22); discharge reason (selected deceased: 15.6% vs excluded deceased: 18.2%; $F_{1,2308}$=2.474; P=.12); ICU admission (6.8% vs 7.3%; $F_{1,2308}$=0.003; P=.96); or admission date (March 27, 2020, ± 8.3 days, vs March 28, 2020, ± 11.6 days). However, there were significant differences in terms of sex (37.2% women vs 42.2% women; $F_{1,2308}$=5.768; P=.02) and days in hospital (mean 9, SD 6, vs mean 8, SD 7; $F_{1,2308}$=4.786; P=.03). Notwithstanding, the effect size was small for both differences (η^2=0.003 and η^2=0.002, respectively).

XSL•FO
RenderX

Table 1. Laboratory tests used to characterize the patients.

Code	Description	Unit
RDW	Red cell distribution width	%
BAS	Basophils	$\times 10^3/\mu L$
BAS%	Percentage of basophils	%
MCHC	Mean corpuscular hemoglobin concentration	g/dL
CREA	Creatinine	mg/dL
EOS	Eosinophils	$\times 10^3/\mu L$
EOS%	Percentage of eosinophils	%
GLU	Glucose	mg/dL
AST	Aspartate transaminase	U/L
ALT	Alanine transaminase	U/L
MCH	Mean corpuscular hemoglobin	pg
HCT	Hematocrit	%
RBC	Red blood cells	$\times 10^6/\mu L$
HB	Hemoglobin	g/dL
K	Potassium	mmol/L
LDH	Lactate dehydrogenase	U/L
LEUC	Leucocytes	$\times 10^3/\mu L$
LYM	Lymphocytes	$\times 10^3/\mu L$
LYM%	Percentage of lymphocytes	%
MONO	Monocytes	$\times 10^3/\mu L$
MONO%	Percentage of monocytes	%
NA	Sodium	mmol/L
NEU	Neutrophils	$\times 10^3/\mu L$
NEU%	Percentage of neutrophils	%
CRP	C-reactive protein	mg/L
PLAT	Platelet count	$\times 10^3/\mu L$
BUN	Blood urea nitrogen	mg/dL
MCV	Mean cell volume	fL
MPV	Mean platelet volume	fL

Clustering

Unsupervised automatic x-means clustering [32]—the implementation in RapidMiner Studio 9.7, Community Edition (RapidMiner, Inc)—was applied to the preprocessed data set that was previously described (see Data Preprocessing section). The algorithm determines the optimum number of clusters so that the intracluster distance of patients is at a minimum, and the intercluster distance of patients is at a maximum. The x-means algorithm was used instead of the more common k-means algorithm to overcome the three major shortcomings of the latter [32]: poor computational scaling, manual selection of the number of clusters, and tendency to local minima. X-means clustering determines the optimal number of clusters by the Bayesian information criterion (BIC), also known as the Schwarz criterion, which is used to maximize the explained variance by the clusters and minimize the number of parameters (k) [32]. X-means clustering is also an improvement over k-means clustering since it tends to create clusters formed by only one sample to minimize inertia [32]. Moreover, the later use of the Davies Bouldin index to evaluate the cluster distributions is also intended to overcome this issue since it considers a mix of both inertia and distortion to quantitatively asses the cluster models (see below). In addition, the automatic selection of the number of clusters by x-means clustering avoids the possible bias in the manual selection of k [32]. This bias is also present in hierarchical agglomerative clustering, where a threshold must be set to obtain the ultimate clusters after the hierarchy is built. Despite the fact that x-means clustering is not completely deterministic, it is certainly very stable with

XSL•FO
RenderX

minimum variations between different runs [32] and is significantly more stable than k-means clustering. However, x-means clustering introduces a bias. Since it uses the BIC to evaluate the cluster models in each iteration, this criterion purposely favors the models with a lower number of clusters. This means that an alternative cluster model with a better Davies Boulding index and a higher number of clusters may have been discarded. However, a higher number of clusters with a better Davies Boulding index usually implies clusters with small numbers of samples—notice that the best index would be obtained by a model of one cluster per sample—which is not desirable at all for the clinical stratification purpose aimed for in this study.

Patients were considered here as vectors with several dimensions equal to the number of variables. In this case, the number of variables taken to apply the clustering algorithm was 842. None of the eight variables about demographics, hospital stay, and outcome measures were included. They were removed from the clustering formation because of the potential ethical controversies and biases (ie, demographics) or prospective information (ie, hospitalization stay and outcome measures). The algorithm was applied using several similarity or distance metrics between patients [33]: the Euclidean distance, the Canberra distance, the Chebyshev distance, the correlation similarity, the cosine similarity, the Dice similarity, the inner product similarity, the Jaccard similarity, the kernel Euclidean distance, the Manhattan distance, the max product similarity, the overlap similarity, the generalized divergence, the Itakura-Saito distance, the Kullback-Leibler divergence, the logarithmic loss, the logistic loss, the Mahalanobis distance, the squared Euclidean distance, and the squared loss. In spite of the fact that we could have had good similarity measure candidates a priori, based on data set characteristics such as dimensionality, the best practice was the selection based on empirical evaluation [34]. To avoid any a priori biases, we empirically tested all measures available in the software and kept the one yielding the best results.

To assess the fitness of the cluster distributions from the algorithm executions with the above metrics, the Davies Bouldin index was calculated for each one of them [35]. The Davies Bouldin index is a common measure that evaluates cluster models [35]. It quantifies the average maximum ratio of the within-cluster scatter to the between-cluster separation for every pair of clusters in a cluster model [35]. In other words, it provides a trade-off between intercluster similarity and intracluster distance [35]. With this definition, the lower the Davies Bouldin index the lower the within-cluster scatter and the higher the between-cluster separation, which is the most desirable property of a cluster model [35]. The Davies Bouldin index allowed us to quantitatively select the best cluster model among those created, one for each similarity measure considered.

Cluster Validation

From the 1457 patients excluded due to missing values (ie, not used to obtain the clusters), we performed a validation analysis with the patients who presented no missing values in the variables that statistically differed between the three clusters obtained. Subsequently, these patients were assigned to one of the clusters previously obtained by using the best distance metric determined in the clustering process described above.

Statistical Analysis

The difference in the 850 variables between all the clusters obtained was tested using a one-way multivariate analysis of variance. Pairwise post hoc comparisons between clusters were analyzed by the Bonferroni test. Significance was accepted at the 5% level (α=.05). The observed power and effect size, as partial η^2, were reported for statistically significant differences.

Results

Table 2 shows the number of clusters and the corresponding David Bouldin index of the cluster distribution of patients obtained by the x-means clustering algorithm for each of the similarity measures tested. Note that the lower the David Bouldin index, the better the cluster distribution (ie, higher intercluster distance and lower intracluster distance). The best cluster distribution (ie, lowest David Bouldin index) was obtained by using the Manhattan distance, which grouped the patients into three clusters.

XSL•FO
RenderX

Table 2. Number of clusters and the corresponding David Bouldin index.

Similarity measure	David Bouldin index	Number of clusters
Euclidean distance	0.948	3
Canberra distance	N/A[a]	1
Chebyshev distance	0.966	3
Correlation similarity	1.400	3
Cosine similarity	1.629	3
Dice similarity	N/A	1
Inner product similarity	N/A	1
Jaccard similarity	1.387	3
Kernel Euclidean distance	1.440	3
Manhattan distance	0.701	3
Max product similarity	N/A	1
Overlap similarity	5.099	4
Generalized divergence	3.445	3
Itakura-Saito distance	5.919	4
Kullback-Leibler divergence	5.677	4
Logarithmic loss	4.595	4
Logistic loss	3.445	3
Mahalanobis distance	4.595	4
Squared Euclidean distance	3.445	3
Squared loss	3.659	3

[a]N/A: not applicable; the David Bouldin index could not be calculated for these measures because they only had one cluster each.

Demographic and clinical characteristics of the patients in the three clusters are shown in Table 3. Notice that this table also shows the values of the eight variables (ie, demographics, hospital stay, and outcome measures) that were not used in the construction of the clusters (marked with a footnote in Table 3). Cluster 1 had a significantly higher proportion of deceased patients (46.6%) than cluster 2 (18.0%) and cluster 3 (10.5%). No significant difference in the percentage of ICU admissions was found between clusters. However, the patients who were admitted to the ICU in cluster 1 stayed a significantly shorter time than patients in cluster 3. No significant difference in sex was found between clusters. Patients in cluster 3 were significantly younger than those in cluster 1. In addition, patients in clusters 1 and 2 presented with a significantly higher heart rate in the emergency department than those in cluster 3. The average oxygen saturation for patients in the emergency department was significantly different between all clusters, whereby patients in cluster 1 had the lowest oxygen saturation and those in cluster 3 had the highest. With respect to previous diseases and surgical procedures, cluster 1 patients presented with a significantly higher percentage of epilepsy and emphysema than those in clusters 2 and 3. In addition, cluster 2 patients presented with a higher percentage of previous surgical procedures, as well as previous thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders, than patients in cluster 3. Cluster 2 patients also presented with a significantly lower percentage of disorders of purine and pyrimidine metabolism than those in clusters 1 and 3. Finally, the percentage of patients who underwent surgical operations during their hospitalization was significantly higher in cluster 1 than in clusters 2 and 3.

Regarding laboratory tests, patients in cluster 1 showed significantly higher levels of serum creatinine, potassium, and blood urea nitrogen than those in clusters 2 and 3; cluster 1 patients also had a significantly higher value of red cell distribution width than did cluster 2 patients. In addition, patients in cluster 2 presented with significantly higher values of lymphocytes and serum levels of sodium, and significantly lower platelet counts than patients in cluster 3. In addition, cluster 3 patients showed lower values of mean corpuscular hemoglobin concentration and leucocytes, serum levels of alanine transaminase (ALT), and percentage of neutrophils than did patients in clusters 1 and 2. Cluster 3 patients had significantly higher values and percentages of eosinophils and percentages of lymphocytes than did patients in clusters 1 and 2. Finally, the laboratory tests that showed significant differences between all clusters were found for the serum levels of aspartate transaminase (AST) (cluster 1 > cluster 2 > cluster 3), lactate dehydrogenase (LDH) (cluster 1 > cluster 2 > cluster 3), C-reactive protein (CRP) (cluster 1 > cluster 2 > cluster 3), and the number of neutrophils (cluster 1 > cluster 2 > cluster 3).

XSL•FO
RenderX

Table 3. Demographic and clinical characteristics of patients (N=853) in the three clusters.

Characteristics	Cluster 1 (n=58)	Cluster 2 (n=300)	Cluster 3 (n=495)	*F* test (*df*=2, 850)	*P* value	η^{2a}	$1-\beta^{b}$
Demographics							
Age (years)[c], mean (SD)	71.1 (13.7)[d]	67.0 (15.1)[d,e]	65.1 (16.2)[e]	3.457	.03	0.009	0.648
Sex (men)[c], n (%)	41 (70.7)[d]	181 (60.3)[d]	313 (63.3)[d]	1.027	.36	0.003	0.23
Hospital stay and outcome measures							
Inpatient hospital days[c], mean (SD)	8.5 (4.9)[d]	8.6 (6.4)[d]	8.3 (5.1)[d]	0.363	.70	0.001	0.109
Discharge reason[c], n (%)				26.054	<.001	0.062	1
Recovered	31 (53.4)[d]	246 (82.0)[e]	443 (89.5)[e]				
Deceased	27 (46.6)[d]	54 (18.0)[e]	52 (10.5)[e]				
Intensive care unit admission[c], n (%)				1.12	.33	0.003	0.248
No	52 (89.7)[d]	277 (92.3)[d]	458 (92.5)[d]				
Yes	6 (10.3)[d]	23 (7.7)[d]	37 (7.5)[d]				
Days until intensive care unit admission[c], mean (SD)	0.2 (0.4)[d]	3.4 (6.3)[d]	2.3 (4.3)[d]	1.393	.26	0.042	0.289
Days in intensive care unit[c], mean (SD)	0.2 (0.4)[d]	4.8 (6.5)[d,e]	7.6 (6.9)[e]	3.747	.03	0.106	0.665
Mechanical ventilation need[c], n (%)	35 (60.3)[d]	177 (59.0)[d]	277 (56.0)[d]	0.163	.85	<0.001	0.075
Vital signs and laboratory tests, mean (SD)							
First heart ratio measurement in the emergency department	98.4 (25.0)[d,e]	100.1 (26.2)[d]	93.5 (24.4)[e]	8.45	<.001	0.021	0.965
First oxygen saturation measurement in the emergency department	84.2 (12.3)[d]	90.1 (7.6)[e]	94.2 (3.6)[f]	81.732	<.001	0.171	1
Last heart ratio measurement in the emergency department	99.0 (25.1)[d,e]	100.1 (26.0)[d]	93.6 (24.7)[e]	8.104	<.001	0.02	0.958
Last oxygen saturation measurement in the emergency department	84.2 (12.2)[d]	90.0 (7.52)[e]	94.2 (3.6)[f]	82.554	<.001	0.172	1
Red cell distribution width (%)	13.6 (1.9)[d]	12.9 (1.84)[e]	13.0 (1.9)[d,e]	3.28	.04	0.008	0.623
Basophils ($\times10^3$/μL)	0.03 (0.03)[d]	0.02 (0.02)[d,e]	0.02 (0.0)[e]	5.545	.004	0.014	0.854
Mean corpuscular hemoglobin concentration (g/dL)	33.9 (1.5)[d]	34.0 (1.17)[d]	33.6 (1.2)[e]	8.602	<.001	0.021	0.968
Creatinine (mg/dL)	1.3 (1.4)[d]	1.0 (0.47)[e]	1.0 (0.5)[e]	9.591	<.001	0.024	0.981
Eosinophils ($\times10^3$/μL)	0.02 (0.04)[d]	0.02 (0.04)[d]	0.04 (0.1)[e]	6.518	.002	0.016	0.908
Eosinophils (%)	0.20 (0.5)[d]	0.3 (0.60)[d]	0.6 (1.2)[e]	10.000	<.001	0.025	0.985
Aspartate transaminase (U/L)	80.3 (48.0)[d]	55.8 (33.4)[e]	32.8 (18.7)[f]	109.193	<.001	0.216	1
Alanine transaminase (U/L)	57.2 (69.1)[d]	50.7 (48.1)[d]	29.5 (23.8)[e]	32.686	<.001	0.076	1
Potassium (mmol/L)	4.6 (0.8)[d]	4.2 (0.6)[e]	4.2 (0.5)[e]	16.957	<.001	0.041	1
Lactate dehydrogenase (U/L)	1339.72 (240.56)[d]	742.5 (122.0)[e]	447.7 (91.5)[f]	1666.635	<.001	0.808	1

XSL•FO
RenderX

Characteristics	Cluster 1 (n=58)	Cluster 2 (n=300)	Cluster 3 (n=495)	*F* test (*df*=2, 850)	*P* value	η^{2a}	$1-\beta^{b}$
Leucocytes ($\times10^3$/μL)	9.9 (4.8)[d]	8.5 (4.2)[d]	6.9 (5.2)[e]	13.055	<.001	0.032	0.997
Lymphocytes ($\times10^3$/μL)	1.0 (0.5)[d,e]	1.0 (0.6)[d]	1.3 (2.1)[e]	3.692	.03	0.009	0.679
Lymphocytes (%)	12.6 (7.8)[d]	14.0 (7.7)[d]	20.0 (9.8)[e]	46.962	<.001	0.106	1
Monocytes (%)	5.1 (2.9)[d]	6.6 (3.9)[d]	8.7 (4.8)[e]	29.321	<.001	0.069	1
Sodium (mmol/L)	136.2 (7.1)[d,e]	136.2 (4.4)[d]	137.2 (4.6)[e]	4.016	.02	0.01	0.718
Neutrophils ($\times10^3$/μL)	8.4 (4.7)[d]	6.9 (4.0)[e]	4.9 (2.7)[f]	45.584	<.001	0.103	1
Neutrophils (%)	81.8 (10.2)[d]	78.8 (9.9)[d]	70.4 (11.9)[e]	62.070	<.001	0.135	1
C-reactive protein (mg/L)	206.1 (131.7)[d]	152.1 (110.0)[e]	64.2 (63.7)[f]	12.930	<.001	0.223	1
Platelet count ($\times10^3$/μL)	229.0 (92.2)[d,e]	236.3 (96.6)[d]	210.3 (87.2)[e]	7.541	.001	0.019	0.944
Blood urea nitrogen (mg/dL)	58.9 (56.6)[d]	41.8 (29.0)[e]	40.5 (29.7)[e]	7.579	.001	0.019	0.945
Diseases and surgical procedures, n (%)							
Previous history of disorders of purine and pyrimidine metabolism	4 (6.9)[d]	4 (1.3)[e]	25 (5.1)[d]	4.179	.02	0.01	0.736
Previous history of epilepsy and recurrent seizures	3 (5.2)[d]	4 (1.3)[e]	2 (0.4)[e]	5.660	.004	0.014	0.862
Previous history of emphysema	3 (5.2)[d]	2 (0.7)[e]	2 (0.4)[e]	6.663	.001	0.017	0.914
Previous history of thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders	0 (0)[d,e]	9 (3.0)[d]	3 (0.6)[e]	4.385	.01	0.011	0.758
Previous history of surgical procedures	0 (0)[d,e]	4 (1.3)[d]	0 (0)[e]	3.753	.02	0.009	0.686
Surgical operations during the current hospitalization	3 (5.2)[d]	2 (0.7)[e]	4 (0.8)[e]	4.880	.008	0.012	0.804

[a]Effect size.

[b]Observed power.

[c]These variables were not used for the cluster construction.

[d-f]Values in the same row, but in different columns, that do not share footnote letters were significantly different after Bonferroni post hoc correction; values in the same row, but in different columns, that share footnote letters were not significantly different.

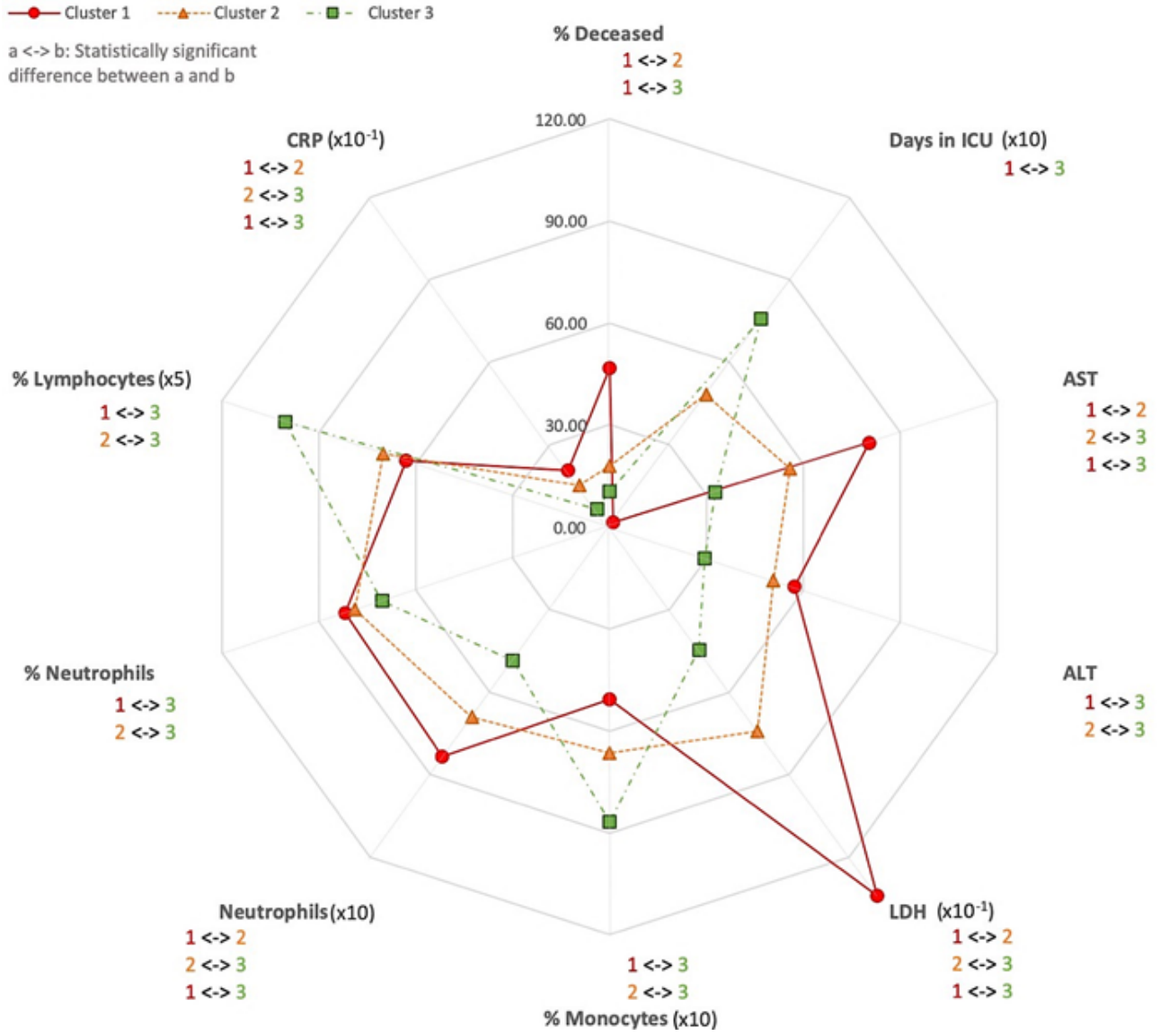
For a clearer characterization of the clusters, Figure 1 shows a radar chart with the variables (ie, hospital stay, outcome measures, and laboratory tests) that showed statistically significant differences among the clusters and a medium or high effect size (η^2>0.06) [36].

A web-based cluster assignment tool, based on the results reported here, can be found online [37].

To test the robustness of the identified clusters, we performed a validation analysis using the initially excluded patients who did not have missing values in the variables that statistically differed among the three clusters (Table 3). Specifically, it was based on six variables (ie, first and last oxygen saturation measurement in the emergency department, AST, LDH, neutrophils, and CRP). For this purpose, we selected 349 patients who were initially excluded and who were assigned to one of the three previously identified clusters by the minimum Manhattan distance to the average values of the six mentioned variables of those clusters. Table 4 shows the differences in demographics, hospital stay, and outcome measures in the three clusters. Indeed, the clusters initially obtained were consistent with the clusters assigned in the validation analysis in terms of age, sex, hospital stay, and outcome measures. Specifically, cluster 1 was the one with the oldest and with the highest proportion of deceased patients. By contrast, cluster 3 was the one with the youngest and with the lowest proportion of deceased patients.

XSL•FO
RenderX

Figure 1. Hospital stay, outcome measures, and laboratory tests that showed statistically significant differences among clusters with a medium or high effect size (η^2>0.06). Note that some variables are scaled (transformation between brackets) for the sake of graph legibility. ALT: alanine transaminase; AST: aspartate transaminase; CRP: C-reactive protein; ICU: intensive care unit; LDH: lactate dehydrogenase.



XSL•FO
RenderX

Table 4. Demographics as well as hospital stay and prognosis of the patients (n=349) selected for the validation analysis in the three clusters.

Characteristics	Cluster 1 (n=18)	Cluster 2 (n=112)	Cluster 3 (n=219)	*F* test (*df*)	*P* value	η^{2a}	$1–\beta^{b}$
Demographics							
Age (years), mean (SD)	72.8 (14.2)[c,d]	71.3 (14.3)[c]	64.2 (15.8)[d]	9.414 (2, 346)	<.001	0.052	0.979
Sex (men), n (%)	14 (77.8)[c]	68 (60.7)[c]	123 (56.2)[c]	1.738 (2, 346)	.18	0.01	0.364
Hospital stay and outcome measures							
Inpatient hospital days, mean (SD)	9.1 (6.4)[c]	9.3 (5.9)[c]	8.0 (5.3)[c]	2.320 (2, 346)	.10	0.013	0.469
Discharge reason, n (%)				22.025 (2, 346)	<.001	0.113	1
Recovered	8 (44.4)[c]	80 (71.4)[d]	200 (91.3)[e]				
Deceased	10 (55.6)[c]	32 (28.6)[d]	19 (8.7)[e]				
Intensive care unit admission, n (%)				4.268 (2, 346)	.02	0.024	0.743
No	16 (88.9)[c,d]	101 (90.2)[c]	213 (97.3)[d]				
Yes	2 (11.1)[c,d]	11 (9.8)[c]	6 (2.7)[d]				
Days until intensive care unit admission, mean (SD)	6.5 (7.8)[c]	4.1 (3.9)[c]	6.3 (13.7)[c]	0.170 (2,16)	.84	0.021	0.072
Days in intensive care unit, mean (SD)	4.5 (0.7)[c]	3.8 (4.5)[c]	3.2 (4.6)[c]	0.082 (2,16)	.92	0.01	0.06
Mechanical ventilation need, n (%)	12 (66.7)[c]	54 (48.2)[c]	96 (43.8)[c]	1.854 (2, 346)	.16	0.011	0.385

[a]Effect size.

[b]Observed power.

[c-e]Values in the same row, but in different columns, that do not share footnote letters were significantly different after Bonferroni post hoc correction; values in the same row, but in different columns, that share footnote letters were not significantly different.

Discussion

With the application of an unsupervised machine learning approach, we could identify and segregate patients with COVID-19 into subgroups depending on the severity of disease, simply by using standard laboratory tests performed during the first medical assessment in the emergency department. We found that inflammatory (ie, CRP), hematologic (ie, number of neutrophils and percentage of monocytes and lymphocytes), and serum biochemical abnormalities (ie, AST, ALT, and LDH), mainly indicating liver dysfunction, detected upon admission to the hospital could predict the severity of the disease. From a sum of 850 variables collected in the emergency department, only four standard laboratory tests (ie, serum levels of AST, LDH, CRP, and the number of neutrophils) were enough to segregate these patients into three separate clusters. Of these, the levels of LDH had the biggest effect size, practically allowing us to differentiate the three clusters linearly. Further, the percentage of monocytes and lymphocytes as well as ALT distinguished cluster 3 patients (ie, less severe) from patients in the other two clusters. Cluster 1 was characterized by the highest proportion of deceased patients; the highest levels of AST, ALT, LDH and CRP; the highest number of neutrophils; and the lowest percentages of monocytes and lymphocytes (Figure 1). Cluster 2 included a lower proportion of deceased patients and intermediate values of the previous laboratory tests (Figure 1). Finally, the lowest proportion of deceased patients; the lowest levels of AST, ALT, LDH and CRP; the lowest number of neutrophils; and the highest percentages of monocytes and lymphocytes characterized cluster 3 (Figure 1).

Our results have several clinical implications. First, age and sex were not considered while building the clusters. Therefore, our unsupervised machine learning approach, based exclusively on the performance of simple laboratory tests at a primordial stage, would permit the establishment of a strategy for rationing of health care resources and to establish a triage protocol, which would support medical decisions in a transparent and ethical way. Second, since the analyzed data are from standard laboratory tests, this method would be especially valuable for underdeveloped and developing regions that lack medical resources and have affordability issues. Finally, we could tailor treatment to each severity group accordingly at a primordial stage (ie, in the emergency department). For example, more aggressive therapies could be considered in patients classified in cluster 1 (ie, the most severe) and not in those in cluster 3 (ie, the least severe).

Initially, SARS-CoV-2 was primarily considered a respiratory pathogen. However, with time, it has behaved like a virus with the potential to cause multisystem involvement [38,39]. Specifically, hepatic injury related to COVID-19 is only beginning to unravel. Elevated liver injury indicators, particularly AST, are strongly associated with a higher mortality risk in patients with COVID-19 [40]. Of note, high serum levels

XSL•FO
RenderX

of LDH predict higher in-hospital mortality in patients with severe and critical condition of COVID-19 [41]. Significant increased CRP levels in the early stages of COVID-19 are correlated with the severity of disease and the degree of internal tissue pathologies [42]. Further, a significant increase in the number of neutrophils with a decrease in the number of lymphocytes, monocytes, and eosinophils may indicate clinical worsening and increased risk of a poor outcome among patients with COVID-19 [43]. Taken together, the presence of elevated biomarkers of inflammation and that of liver injury in serum, as well as the number of neutrophils at admission, are indicative of multiple organ failure in patients with COVID-19 that could lead to death. Our laboratory findings are in agreement with other previous studies worldwide [44-46].

Although one previous multicenter study, based on the analyses of demographics, comorbidities, vital signs, and laboratory test results upon admission, that evaluated the prediction of disease course in patients with COVID-19 has been undertaken [18], there remains much to learn about applying machine learning techniques regarding this novel infectious disease. Comparison with that study is difficult, as they had used different variables and techniques. The accuracy of the model could be influenced by several factors, including the methods. Feature extraction methods, feature selection or classification tools, number of subjects, and demographics are also important considerations. Besides, most COVID-19 diagnostics and prognostic models that have evolved to date have a high risk of generating bias leading to inequality [47], mainly due to the high influence of demographic variables, especially age and sex, in those models and to the nonblinded nature of the supervised machine learning approach between predictors and outcome measures. In fact, our results confirmed that age and sex had a similar and low discriminant value to separate the three clusters (Table 3). Nevertheless, the results obtained in our study are in line with most previous work based on supervised machine learning techniques in COVID-19 [18,47].

The study should be interpreted within the context of several limitations. First, the patients in this study may represent a selected group of patients with COVID-19 (ie, patients with a more severe disease, since all of them were admitted to the hospital); hence, it is questionable as to what extent our results could be generalized to the entire population of patients with COVID-19. The reason for this was that the extreme circumstances in our hospitals at the peak of this pandemic permitted the hospitalization of only the most severe cases. Notwithstanding, our aim was to detect severity subgroups among patients with COVID-19 upon admission to the hospital. Second, we only kept the records (ie, patients), laboratory tests, and clinical variables from 853 patients from the data set due to the high number of missing values in the remaining 1457 patients. Despite this, the results have been robust.

In closing, to the authors' knowledge, the work presented in this paper is the first attempt to use unsupervised machine learning to identify severity subgroups among patients with COVID-19 upon admission. A few affordable, simple, and standard laboratory tests, which are expected to be available in any emergency department, have shown promising discriminative power for characterization of severity subgroups among patients with COVID-19. We have also provided an online severity cluster assignment tool for patients with COVID-19 who are admitted to the emergency department [37]. This could permit the classification of patients according to severity subgroups and, hence, initiate earlier interventions.

Authors' Contributions

JBL collaborated in the conception, organization, and execution of the research project; the writing of the first draft of the manuscript; and the review and critique of the manuscript. MDDC collaborated in the conception and organization of the research project, the statistical analyses, writing of the first draft of the manuscript, and the review and critique of the manuscript. AE collaborated in the organization of the research project and the review and critique of the manuscript. RG collaborated in the organization of the research project and the review and critique of the manuscript. SD collaborated in the organization of the research project and the review and critique of the manuscript. JIS collaborated in the conception and organization of the research project, the statistical analyses, the writing of the first draft of the manuscript, and the review and critique of the manuscript.

Conflicts of Interest

None declared.

References

1. Joebges S, Biller-Andorno N. Ethics guidelines on COVID-19 triage-an emerging international consensus. Crit Care 2020 May 06;24(1):201 [FREE Full text] [doi: 10.1186/s13054-020-02927-1] [Medline: 32375855]
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
3. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artif Intell Healthc 2020:25-60. [doi: 10.1016/b978-0-12-818438-7.00002-2]
4. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 2020 Apr;104:101822 [FREE Full text] [doi: 10.1016/j.artmed.2020.101822] [Medline: 32499001]
5. Serrano JI, Romero JP, Castillo MDD, Rocon E, Louis ED, Benito-León J. A data mining approach using cortical thickness for diagnosis and characterization of essential tremor. Sci Rep 2017 May 19;7(1):2190 [FREE Full text] [doi: 10.1038/s41598-017-02122-3] [Medline: 28526878]

XSL•FO
RenderX

6. Benito-León J, Louis ED, Mato-Abad V, Sánchez-Ferro A, Romero JP, Matarazzo M, et al. A data mining approach for classification of orthostatic and essential tremor based on MRI-derived brain volume and cortical thickness. Ann Clin Transl Neurol 2019 Dec;6(12):2531-2543 [FREE Full text] [doi: 10.1002/acn3.50947] [Medline: 31769622]
7. Mato-Abad V, Labiano-Fontcuberta A, Rodríguez-Yáñez S, García-Vázquez R, Munteanu CR, Andrade-Garda J, et al. Classification of radiologically isolated syndrome and clinically isolated syndrome with machine-learning techniques. Eur J Neurol 2019 Jul;26(7):1000-1005. [doi: 10.1111/ene.13923] [Medline: 30714276]
8. Michalski RS, Bratko I, Kubat M. Machine Learning and Data Mining: Methods and Applications. Chichester, UK: John Wiley & Sons, Ltd; 1998.
9. Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. J Crit Care 2018 Jun;45:156-162. [doi: 10.1016/j.jcrc.2018.02.010] [Medline: 29486341]
10. Alsuliman T, Humaidan D, Sliman L. Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality? Curr Res Transl Med 2020 Nov;68(4):245-251. [doi: 10.1016/j.retram.2020.01.002] [Medline: 32029403]
11. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. J Biomed Inform 2018 Sep;85:30-39 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.004] [Medline: 30016722]
12. Shang Y, Liu T, Wei Y, Li J, Shao L, Liu M, et al. Scoring systems for predicting mortality for severe patients with COVID-19. EClinicalMedicine 2020 Jul;24:100426 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100426] [Medline: 32766541]
13. Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. Environ Res 2020 Sep;188:109890 [FREE Full text] [doi: 10.1016/j.envres.2020.109890] [Medline: 32846654]
14. Fu J, Huang P, Zhang S, Yao Q, Han R, Liu H, et al. The value of serum amyloid A for predicting the severity and recovery of COVID-19. Exp Ther Med 2020 Oct;20(4):3571-3577 [FREE Full text] [doi: 10.3892/etm.2020.9114] [Medline: 32855710]
15. Hou H, Zhang B, Huang H, Luo Y, Wu S, Tang G, et al. Using IL-2R/lymphocytes for predicting the clinical progression of patients with COVID-19. Clin Exp Immunol 2020 Jul;201(1):76-84 [FREE Full text] [doi: 10.1111/cei.13450] [Medline: 32365221]
16. Zeng Z, Yu H, Chen H, Qi W, Chen L, Chen G, et al. Longitudinal changes of inflammatory parameters and their correlation with disease severity and outcomes in patients with COVID-19 from Wuhan, China. Crit Care 2020 Aug 27;24(1):525 [FREE Full text] [doi: 10.1186/s13054-020-03255-0] [Medline: 32854750]
17. Zhang C, Qin L, Li K, Wang Q, Zhao Y, Xu B, et al. A novel scoring system for prediction of disease severity in COVID-19. Front Cell Infect Microbiol 2020;10:318 [FREE Full text] [doi: 10.3389/fcimb.2020.00318] [Medline: 32582575]
18. Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B, Hellenic COVID-19 Study Group. COVID-19 mortality risk assessment: An international multi-center study. PLoS One 2020;15(12):e0243262 [FREE Full text] [doi: 10.1371/journal.pone.0243262] [Medline: 33296405]
19. Ye W, Lu W, Tang Y, Chen G, Li X, Ji C, et al. Identification of COVID-19 clinical phenotypes by principal component analysis-based cluster analysis. Front Med (Lausanne) 2020;7:570614 [FREE Full text] [doi: 10.3389/fmed.2020.570614] [Medline: 33282887]
20. COVID-19 decision support tool. NHS England. URL: http://prod-upp-image-read.ft.com/765d3430-7a57-11ea-af44-daa3def9ae03 [accessed 2021-03-22]
21. COVID-19 rapid guideline: Critical care in adults. National Institute for Health and Care Excellence (NICE). 2020 Mar 20. URL: https://www.nice.org.uk/guidance/ng159 [accessed 2021-03-22]
22. Ahadi S, Zhou W, Schüssler-Fiorenza Rose SM, Sailani MR, Contrepois K, Avina M, et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. Nat Med 2020 Jan;26(1):83-90 [FREE Full text] [doi: 10.1038/s41591-019-0719-5] [Medline: 31932806]
23. Marengoni A, Zucchelli A, Vetrano DL, Armellini A, Botteri E, Nicosia F, et al. Beyond chronological age: Frailty and multimorbidity predict in-hospital mortality in patients with coronavirus disease 2019. J Gerontol A Biol Sci Med Sci 2021 Feb 25;76(3):e38-e45 [FREE Full text] [doi: 10.1093/gerona/glaa291] [Medline: 33216846]
24. Gómez-Moreno C, Hernández-Ruiz V, Hernández-Gilsoul T, Avila-Funes JA. Clinical decision making in older adults with COVID-19 in developing countries: Looking beyond chronological age. Rev Invest Clin 2020;72(3):127-134. [doi: 10.24875/RIC.20000131] [Medline: 32584321]
25. Ehni H, Wahl H. Six propositions against ageism in the COVID-19 pandemic. J Aging Soc Policy 2020;32(4-5):515-525. [doi: 10.1080/08959420.2020.1770032] [Medline: 32491963]
26. Perrotta F, Corbi G, Mazzeo G, Boccia M, Aronne L, D'Agnano V, et al. COVID-19 and the elderly: Insights into pathogenesis and clinical decision-making. Aging Clin Exp Res 2020 Aug;32(8):1599-1608 [FREE Full text] [doi: 10.1007/s40520-020-01631-y] [Medline: 32557332]

XSL•FO
RenderX

27. Juul S, Nielsen EE, Feinberg J, Siddiqui F, Jørgensen CK, Barot E, et al. Interventions for treatment of COVID-19: Second edition of a living systematic review with meta-analyses and trial sequential analyses (The LIVING Project). PLoS One 2021;16(3):e0248132 [FREE Full text] [doi: 10.1371/journal.pone.0248132] [Medline: 33705495]
28. Rahman A, Jahan Y. Defining a 'risk group' and ageism in the era of COVID-19. J Loss Trauma 2020 Apr 25;25(8):631-634. [doi: 10.1080/15325024.2020.1757993]
29. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020;3:81 [FREE Full text] [doi: 10.1038/s41746-020-0288-5] [Medline: 32529043]
30. Chang E, Kannoth S, Levy S, Wang S, Lee JE, Levy BR. Global reach of ageism on older persons' health: A systematic review. PLoS One 2020;15(1):e0220857 [FREE Full text] [doi: 10.1371/journal.pone.0220857] [Medline: 31940338]
31. Covid Data Save Lives. HM Hospitales. URL: https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version [accessed 2020-05-14]
32. Pelleg D, Moore A. X-means: Extending k-means with efficient estimation of the number of clusters. In: Proceedings of the 17th International Conference on Machine Learning. 2000 Presented at: 17th International Conference on Machine Learning; June 29-July 2, 2000; Stanford, CA p. 727-734 URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B8BE3002265BD61591083E9412FB2C9B?doi=10.1.1.19.3377&rep=rep1&type=pdf
33. Cha SH. Comprehensive survey on distance/similarity measures between probability density functions. Int J Math Models Methods Appl Sci 2007;1(4):300-307 [FREE Full text]
34. Shirkhorshidi AS, Aghabozorgi S, Wah TY. A comparison study on similarity and dissimilarity measures in clustering continuous data. PLoS One 2015;10(12):e0144059 [FREE Full text] [doi: 10.1371/journal.pone.0144059] [Medline: 26658987]
35. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. J Intell Inf Syst 2001;17(2):107-145. [doi: 10.1023/a:1012801612483]
36. Vacha-Haase T, Thompson B. How to estimate and interpret various effect sizes. J Couns Psychol 2004;51(4):473-481. [doi: 10.1037/0022-0167.51.4.473]
37. COVID-19 clinical severity group assessment. Neural and Cognitive Engineering Group, Automation and Robotics Center (CAR), CSIC-UPM. 2021. URL: https://g-nec.car.upm-csic.es/COVID19-severity-group-assessment/ [accessed 2021-05-02]
38. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
39. Roy D, Ghosh R, Dubey S, Dubey MJ, Benito-León J, Kanti Ray B. Neurological and neuropsychiatric impacts of COVID-19 pandemic. Can J Neurol Sci 2021 Jan;48(1):9-24 [FREE Full text] [doi: 10.1017/cjn.2020.173] [Medline: 32753076]
40. Lei F, Liu Y, Zhou F, Qin J, Zhang P, Zhu L, et al. Longitudinal association between markers of liver injury and mortality in COVID-19 in China. Hepatology 2020 Aug;72(2):389-398 [FREE Full text] [doi: 10.1002/hep.31301] [Medline: 32359177]
41. Dong X, Sun L, Li Y. Prognostic value of lactate dehydrogenase for in-hospital mortality in severe and critically ill patients with COVID-19. Int J Med Sci 2020;17(14):2225-2231 [FREE Full text] [doi: 10.7150/ijms.47604] [Medline: 32922185]
42. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, et al. C-reactive protein correlates with computed tomographic findings and predicts severe COVID-19 early. J Med Virol 2020 Jul;92(7):856-862 [FREE Full text] [doi: 10.1002/jmv.25871] [Medline: 32281668]
43. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): A meta-analysis. Clin Chem Lab Med 2020 Jun 25;58(7):1021-1028. [doi: 10.1515/cclm-2020-0369] [Medline: 32286245]
44. Poggiali E, Zaino D, Immovilli P, Rovero L, Losi G, Dacrema A, et al. Lactate dehydrogenase and C-reactive protein as predictors of respiratory failure in CoVID-19 patients. Clin Chim Acta 2020 Oct;509:135-138 [FREE Full text] [doi: 10.1016/j.cca.2020.06.012] [Medline: 32531257]
45. Pourbagheri-Sigaroodi A, Bashash D, Fateh F, Abolghasemi H. Laboratory findings in COVID-19 diagnosis and prognosis. Clin Chim Acta 2020 Nov;510:475-482 [FREE Full text] [doi: 10.1016/j.cca.2020.08.019] [Medline: 32798514]
46. Soraya GV, Ulhaq ZS. Crucial laboratory parameters in COVID-19 diagnosis and prognosis: An updated meta-analysis. Med Clin (Engl Ed) 2020 Aug 28;155(4):143-151 [FREE Full text] [doi: 10.1016/j.medcle.2020.05.004] [Medline: 32864456]
47. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]

Abbreviations

ALT: alanine transaminase
AST: aspartate transaminase
BIC: Bayesian information criterion

XSL•FO
RenderX

CRP: C-reactive protein
ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision
ICU: intensive care unit
LDH: lactate dehydrogenase

Edited by C Basch; submitted 23.11.20; peer-reviewed by AV Das, X Cheng; comments to author 01.03.21; revised version received 12.03.21; accepted 25.03.21; published 27.05.21.

Please cite as:
Benito-León J, del Castillo MD, Estirado A, Ghosh R, Dubey S, Serrano JI
Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups Among Patients with COVID-19: Observational Longitudinal Study
J Med Internet Res 2021;23(5):e25988
URL: https://www.jmir.org/2021/5/e25988
doi:10.2196/25988
PMID:33872186

©Julián Benito-León, Mª Dolores del Castillo, Alberto Estirado, Ritwik Ghosh, Souvik Dubey, J Ignacio Serrano. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2021. 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

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:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e32120

Abstract

Background: Persistent olfactory dysfunction is a significant complication of SARS-CoV-2 infection. Olfactory training involving aromatic oils has been recommended to improve olfactory recovery, but quantitative data are missing.

Objective: We aimed to quantify the benefit of olfactory training and visual stimulation assisted by a dedicated web application for patients who experienced olfactory dysfunction for ≥1 month.

Methods: We performed an observational, real-life, data-based study on a cohort of patients who experienced at least 1 month of persistent olfactory dysfunction between January 30 and March 26, 2021. An analysis was performed after a mean olfactory training time of 4 weeks, and at least 500 patients were assessable for primary outcome assessment. Participants exposed themselves twice daily to odors from 4 high-concentration oils and visual stimulation assisted by a dedicated web application. Improvement was defined as a 2-point increase on a 10-point, self-assessed olfactory visual analogue scale.

Results: In total, 548 patients were assessable for primary outcome assessment. The mean baseline, self-assessed olfactory score was 1.9 (SD 1.7), and this increased to 4.6 (SD 2.8) after a mean olfactory training time of 27.7 days (SD 17.2). Olfactory training was associated with at least a 2-point increase in 64.2% (352/548) of patients. The rate of patients' olfactory improvement was higher for patients who trained for more than 28 days than that rate for patients who trained for less than 28 days (73.3% vs 59%; P=.002). The time to olfactory improvement was 8 days faster for patients with hyposmia compared to the time to improvement for patients with anosmia (P<.001). This benefit was observed regardless of the duration of the olfactory dysfunction.

Conclusions: Olfactory training and visual stimulation assisted by a dedicated web application was associated with significant improvement in olfaction, especially after 28 days of olfactory training.

(J Med Internet Res 2021;23(5):e29583) doi:10.2196/29583

XSL•FO
RenderX

KEYWORDS

olfactory dysfunction; SARS-CoV-2; olfactory training; web application, eHealth; real-life study; COVID-19; app; training; stimulation; olfactory; dysfunction; smell; observational; senses; nose

Introduction

Anosmia is a frequent symptom of SARS-CoV-2 infection, and its duration is usually less than 2 weeks before recovery [1-3]. However, at least 10% of patients with SARS-CoV-2 infection will experience persistent and chronic olfactory dysfunction such as diminished smell (hyposmia) or the loss of smell (anosmia), which have been shown to result in a decreased quality of life, depressive symptoms, and nutrition issues [4-6]. One treatment option that is recommended for persistent olfactory dysfunction is daily olfactory training involving high-concentration aromatic oils [7]. This showed significant results in treating postinfectious olfactory loss in a randomized, controlled, multicenter study [8]. In this trial, after 18 weeks of olfactory training, olfactory function improved in 63% of patients who experienced olfactory dysfunction for a duration of less than 12 months and used high-concentration oils, whereas olfactory function improved in 19% of patients in the control group who used low-concentration oils. Moreover, the combination of visual stimulations and olfactory training may improve recovery results [9].

No data about olfactory training for persistent olfactory dysfunction are available on patients with SARS-CoV-2 infection and persistent olfactory dysfunction, but most patients who experience hyposmia or anosmia for 30 days or more seem to have a low rate of spontaneous recovery [4].

In order to quantitatively study the time course of olfactory scores during olfactory training in real life, we developed a web application dedicated to olfactory training and visual stimulations as well as the self-assessment and follow-up of olfactory scores. We assessed the results in a real-life observational study.

Methods

The web application users were recruited via a national media campaign in France that was disseminated through social media, radio, and magazines between January 30 and February 15, 2021.

This observational, data-based study was approved by the French National Health Data Institute, which reviews the ethical conduct of human subjects research, data confidentiality, and safety. To participate, individuals were required to connect to the free covidanosmia.eu web application and provide electronic agreement. Respondents anonymously self-entered sociodemographic data and real-time polymerase chain reaction test results and confirmed a diagnosis of SARS-CoV-2–related olfactory dysfunctions. Patients were also asked to complete items about comorbidities, the duration of olfactory symptoms, and the self-assessed intensity of olfactory dysfunction by providing subjective ratings with a visual analogue scale of 0 (no smell) to 10 (no smell alteration) [10]. Patients were retained in the study analysis if they were diagnosed with SARS-CoV-2–related olfactory dysfunction that persisted for at least 1 month and reported at least 7 days of olfactory training, and if their last olfactory function assessment on the web application diary was available. The exclusion criteria were normosmia (visual scale score of >7); other causes of olfactory dysfunction such as chronic rhinosinusitis, nasal polyposis, allergic or idiopathic rhinitis, posttraumatic olfactory loss, and other acute or chronic nasal diseases (eg, acute viral infections); malignant tumors or oncology therapies (radiation therapy and chemotherapy), and a history of surgery for the nose or paranasal sinuses.

Patients had to obtain the olfactory training kit from the web application or from their pharmacist. Olfactory training was performed for a maximum period of 16 weeks. The web application provides videos, tutorials for the training, and periodic encouragements. Participants exposed themselves twice daily to odors from the following four high-concentration oils: phenyl ethyl alcohol (rose odor from *Geranium rosa*), eucalyptol (eucalyptus odor), citronellal (lemon odor), and eugenol (cloves odor). These four odorants were chosen to represent the primary odor categories created by Henning [11,12]. Participants sniffed each odor for approximately 15 seconds while blinded and repeated this process 30 seconds later once while the name and picture of the oil component was on the screen of the web application (eg, a picture of a lemon during the lemon oil sniffing process). Patients were asked to train in the morning and in the evening, resulting in a total of 4 exposures per day per odor. They were asked to keep a daily diary on the web application, in which they rated their overall olfactory abilities for each oil with subjective ratings on a visual analogue scale.

We assessed the rate of self-assessed improvement in overall olfactory function along with training times by using data that were collected anonymously from the web application diaries of patients. Improvement was defined as an increase of ≥2 points on the olfactory visual analogue scale. The analysis was performed when the mean olfactory training time of the study population was at least 4 weeks and when at least 500 patients were assessable for primary outcome assessment.

Categorical variables were summarized by using frequencies and percentages, and chi-square tests or Fisher exact tests were used to make comparisons. For quantitative variables, which were summarized with descriptive statistics, the following values were presented: N values, means, and SDs. A *t* test was used to compare groups, and the analysis of variance test was used for comparisons of more than 2 groups.

The Kaplan-Meier methodology was used to summarize time-to-event variables. Plots of Kaplan-Meier product limit estimates for time-to-event variables were drawn, and medians were presented in addition to CIs set at 95%. To compare Kaplan-Meier curves of the two groups, the log-rank test was used.

XSL•FO
RenderX

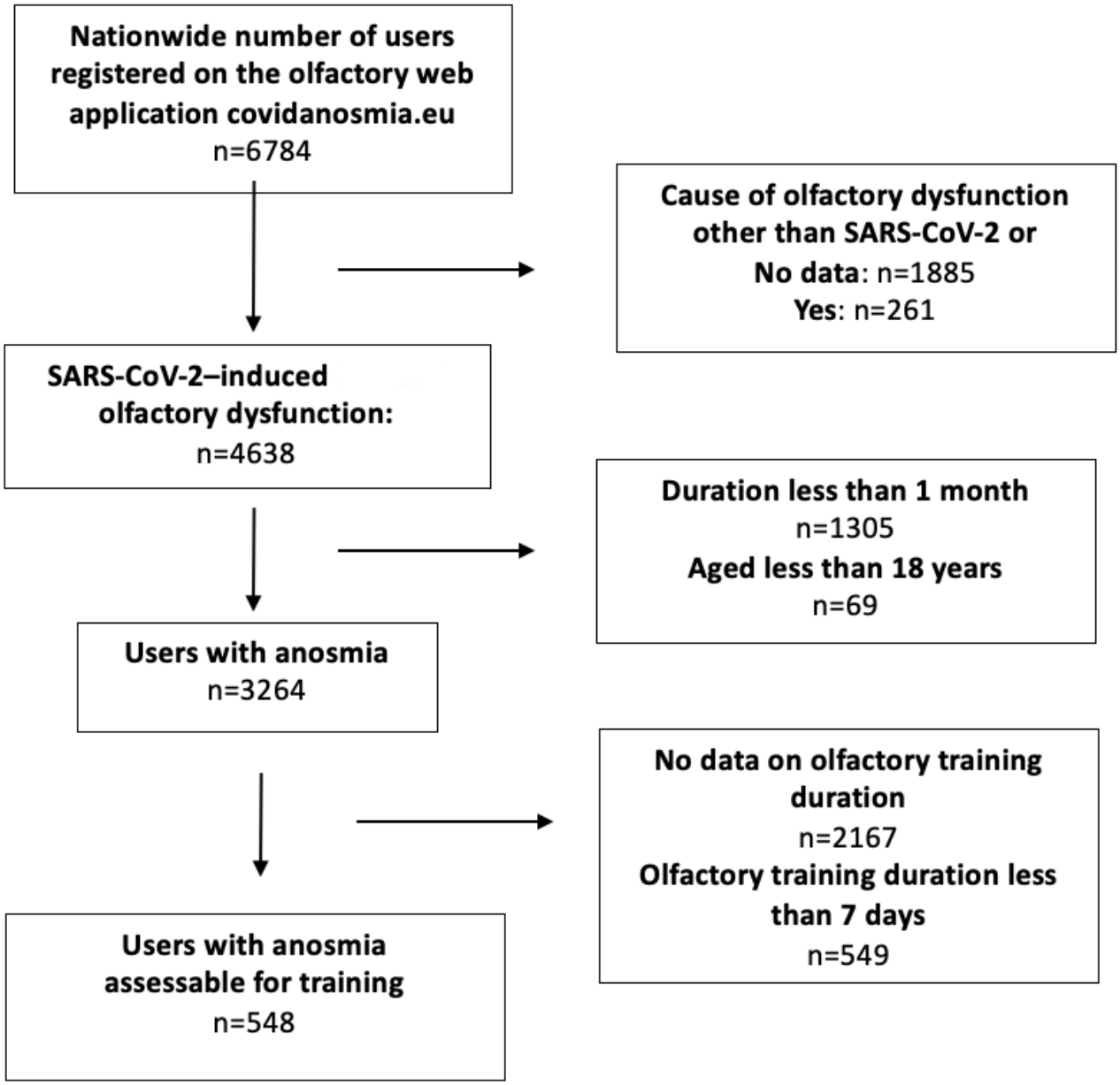
The level of statistical significance was 5% for all statistical tests (exploratory tests). To analyze predictive factors of assessment, logistic regression was used in order to calculate odds ratios, which were presented with CIs set at 95%.

All statistical analyses were conducted with SAS (Statistical Analysis System), version 9.3 (SAS Institute Incorporated).

Results

Between January 30 and March 26, 2021, the web application was used by 6755 unique individuals who completed the baseline questionnaires. Of these individuals, 548 met the inclusion criteria and were assessable for outcome assessments (Figure 1).

Figure 1. Flowchart of patients who used a web application for olfactory training.



The 548 assessable patients' median age was 42 years (range 18-84 years). Of these patients, 65.5% (n=359) were female, 32.1% (n=176) estimated having a smell sense that was more developed than average before olfactory dysfunction, 69.3% (n=380) reported that smell sense had an important role in their life, and 30.7% (n=168) did not care about their smell sense before SARS-CoV-2 infection. Of the 548 assessable patients, 111 (20.3%) experienced anosmia (level 0 on the olfactory scale), 287 (52.4%) experienced severe hyposmia (level 1 or 2 on the olfactory scale), 125 (22.8%) experienced moderate hyposmia (levels 3-5 on the olfactory scale), 25 (4.6%) experienced mild hyposmia (levels 6-7), 279 (50.9%) reported a reduction in or loss of taste, and 289 (52.7%) reported parosmia. Patients' baseline characteristics are shown in Table 1. The mean baseline olfactory function of users who were registered on the web application but underwent less than 7 days olfactory training or did not record their last olfactory function assessment on the web application diary (n=2824; olfactory function score: mean 2.23) was higher than that of the studied population (n=548 patients; olfactory function score: mean 1.9; Student test $P<.001$).

XSL•FO
RenderX

The mean baseline, self-assessed olfactory score was 1.9 (SD 1.7), and this increased to 4.6 (SD 2.8) after a mean olfactory training time of 27.7 days (SD 17.2 days; range 7-65 days).

Olfactory training was associated with at least a 1-point increase on the olfactory scale in 82.1% (450/548) of patients, at least a 2-point increase (ie, the primary outcome) in 64.2% (352/548) of patients, and at least a 3-point increase in 49.3% (270/548) of patients during the study period. The rate of olfactory improvement in patients who experienced anterior olfactory dysfunction for 12 months was 58.3%. With regard to patients whose olfactory function score increased by at least 2 points, their scores increased by a mean of 4.1 points (SD 1.9 points).

Table 1. Patients' characteristics.

Variables	Value, n (%)
Sex	
Male	189 (34.5)
Female	359 (65.5)
Smell level before smell dysfunction	
Standard	354 (64.6)
Less developed than average	18 (3.3)
More developed than average	176 (32.1)
Role of smell before smell loss	
Did not care about smell	168 (30.7)
Important role	380 (69.3)
Smell level at baseline	
0	111 (20.3)
1-2	287 (52.4)
3-5	125 (22.8)
6-7	25 (4.6)
Olfactory dysfunction duration (months)	
1-2	61 (11.1)
2.1-3	250 (45.6)
3.1-6	167 (30.5)
6.1 to ≥12	70 (12.8)
Taste dysfunction	
None	269 (49.1)
Dysfunction	279 (50.9)
Parosmia	
No	259 (47.3)
Yes	289 (52.7)

The duration of the training was associated with better outcomes, and the time to olfactory function improvement was longer in patients with anosmia (olfactory training duration for a 50% probability of improvement: mean 41 days; range 36-53 days) than in patients with hyposmia (mean 33 days; range 28-36 days; log-rank $P<.001$; Figure 2). There were no significant differences in the duration of training among patients with severe, moderate, and mild hyposmia (severe vs moderate $P=.052$; severe vs mild $P=.96$ and moderate vs mild $P=.87$).

The rate of patients' olfactory improvement (at least a 2-point increase on the olfactory scale) was higher for patients who trained for more than 28 days than that rate for patients who trained for less than 28 days (73.3% vs 59%; $P=.002$). Patients who underwent 28 days of olfactory training or more and benefited from olfactory improvement exhibited a mean improvement of 4.4 points (SD 2.0 points) on the olfactory scale, whereas a mean improvement of 3.8 points (SD 1.8) was observed in patients who underwent less than 28 days olfactory training (Student test $P=.01$).

The mean improvement in self-assessed olfactory scale scores was similar regardless of the anteriority of the olfactory dysfunction ($P=.70$; Figure 3). No other predictive factors were highlighted (Table 2).

XSL•FO
RenderX

Figure 2. Cumulative incidence of olfactory improvement according to the olfactory training durations of patients with anosmia and hyposmia. The data of patients with mild, moderate, and severe hyposmia were pooled in the blue curve.

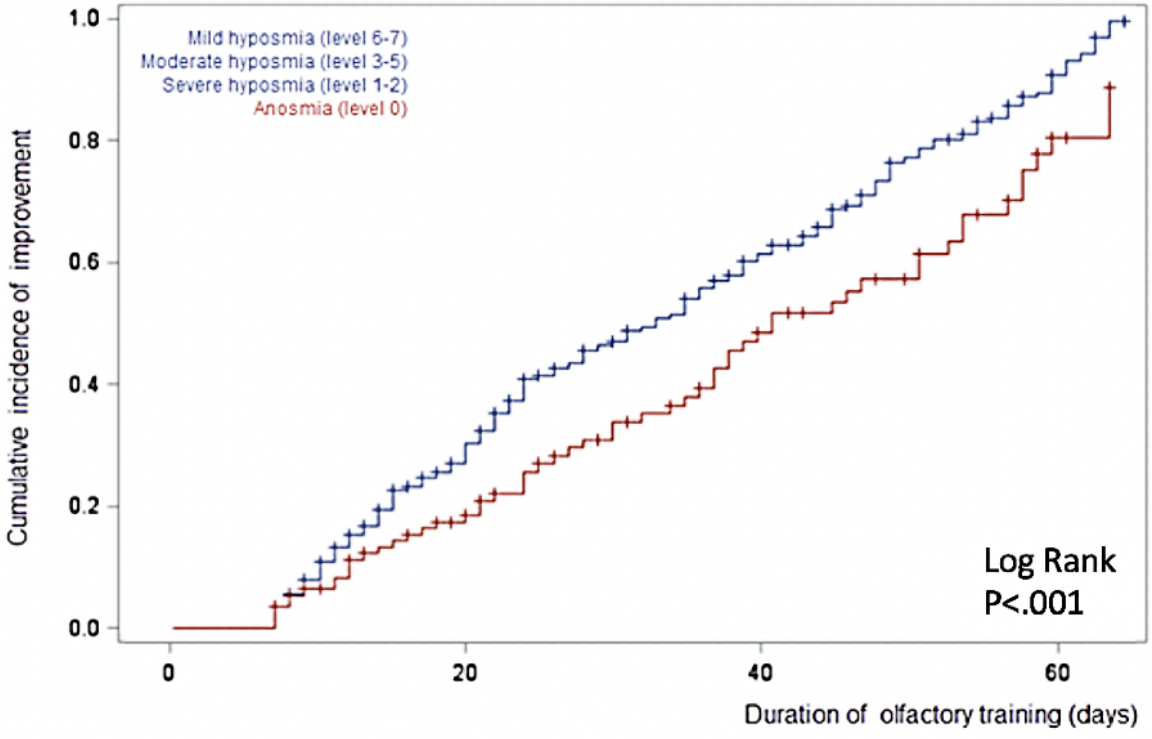
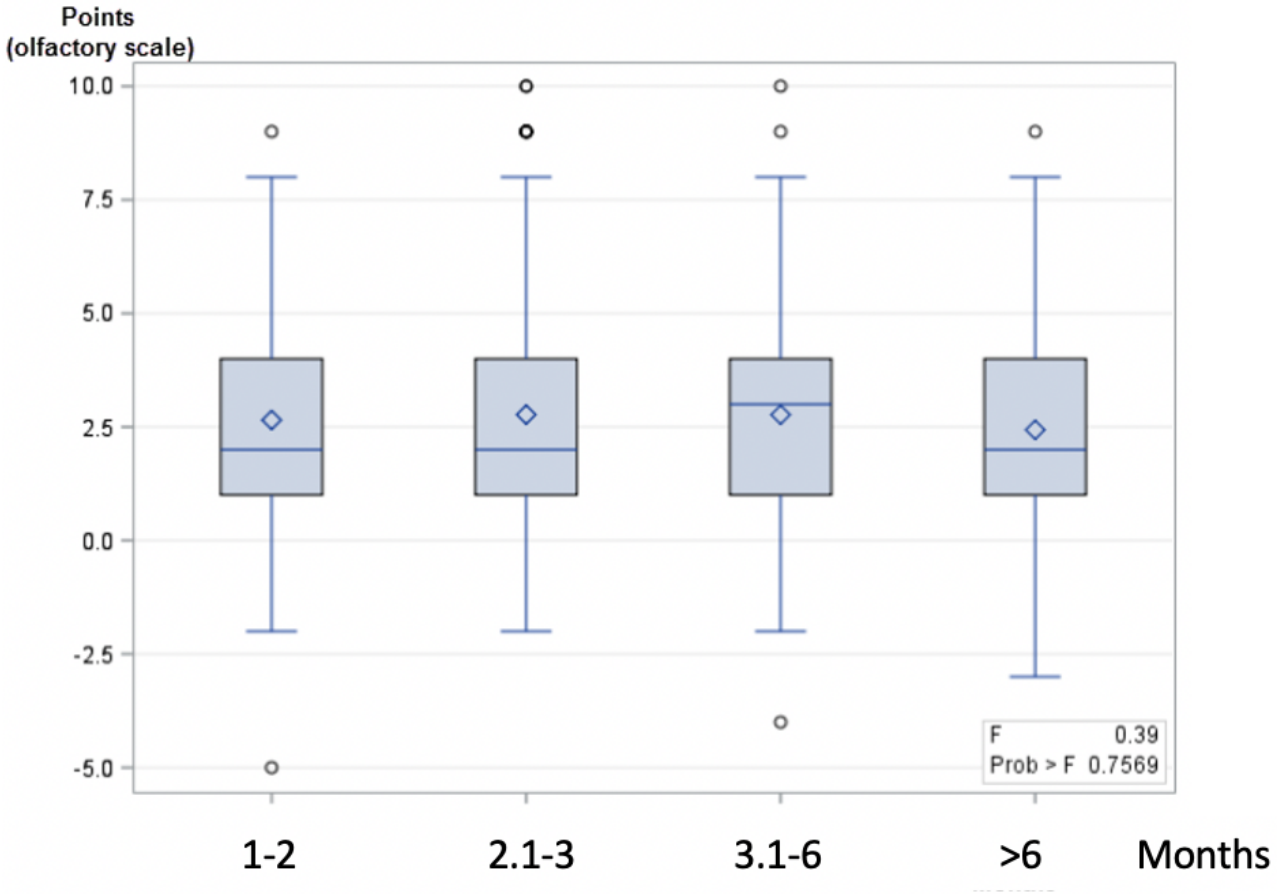


Figure 3. Mean improvement of olfactory function stratified by the duration of persistent olfactory dysfunction. Improvement was assessed with a self-assessed olfactory scale of 0-10 after olfactory training.



XSL•FO
RenderX

Table 2. Logistic regression analysis for determining the predictive factors of olfactory function improvement (ie, an increase of ≥2 points on the olfactory scale).

Variables	Univariate analysis	
	Hazard ratio (95% CI)	*P* value
Gender		.77
Male	1 (N/A[a])	
Female	0.945 (0.654-1.366)	
Age	1.012 (0.997-1.026)	.11
COVID-19 tests		>.99
Positive test	1 (N/A)	
No test	0.934 (0.220-3.955)	
Yes; negative test	1.058 (0.462-2.425)	
Yes; no test	1.074 (0.521-2.212)	
Anosmia duration (months)		.85
>6	1 (N/A)	
1	1.287 (0.627-2.643)	
2	1.078 (0.625-1.861)	
3-6	1.212 (0.680-2.160)	
Smell importance		.65
Standard	1 (N/A)	
Less than standard	0.666 (0.256-1.730)	
Better than standard	0.909 (0.624-1.325)	
Taste lost		.50
No	1 (N/A)	
Yes	1.128 (0.795-1.600)	
Parosmia		.44
No	1 (N/A)	
Yes	1.149 (0.810-1.630)	
Training duration (days)		.002
<28	1 (N/A)	
≥28	1.802 (1.247-2.604)	

[a]N/A: not applicable.

After a mean olfactory training time of 28 days, we observed that 17.1% (94/548) of patients had an olfactory score of ≥8. This high recovery was observed in patients regardless of the anteriority of the olfactory dysfunction (*P*=.93); 43.6% (41/94) of these patients experienced more than 3 months of olfactory dysfunction whereas 56.4% (53/94) experienced less than 3 months of dysfunction. However, patients with an olfactory training score of ≥8 had significantly higher baseline scores than users who did not achieve this score after training; the mean baseline scores were 2.9 (SD 2.0) for patients who achieved high recovery and 1.6 (SD 1.5) for those who achieved lower recovery (*P*<.001).

Discussion

This study is the first to prospectively assess the real-life benefit of olfactory training for patients who experience persistent olfactory dysfunction after SARS-CoV-2 infection. The mean duration of training was 28 days. In our cohort of 548 patients who underwent olfactory training assisted by the web application, an improvement of 2 points or more in a subjective self-assessed olfactory scale was reported by 64.2% (352/548) of patients. Beyond 28 days of training, the rate of improvement was significantly higher than the rate of improvement after <28 days of training (72.2% vs 59%; *P*=.002). The time to olfactory improvement was 8 days longer for patients with anosmia compared to the time to improvement for patients with hyposmia. Improvement was observed regardless of the

XSL•FO
RenderX

anteriority of the olfactory dysfunction. High recovery, that is, normal or subnormal self-assessed olfactory function, was observed in 17% (93/548) of patients regardless of the anteriority of the olfactory dysfunction, but high recovery occurred more frequently in patients with higher baseline olfactory scores (mean 2.9).

Our data are in line with a previous randomized trial on postinfectious olfactory loss [8]. In this trial, after 18 weeks of olfactory training, olfactory function improved in 63% of patients who experienced olfactory dysfunction for a duration of less than 12 months and used high-concentration oils, whereas olfactory function improved in 19% of patients in the control group who used low-concentration oils. We used high-concentration oils in combination—the same 4 odorants used by Damm et al [8]—in a 2-step process for each oil. The first step involved blind olfactory stimulation with a given oil like in the Damm et al [8] trial. The second step, which followed the first step and involved the same oil, was enriched via visual stimulation with a picture of the oil component on a smartphone screen, which was delivered by the web application. We chose this new approach to reinforce the olfactory trial with a mixed olfactory-visual trial, as some previous data have suggested that human olfactory perception can substantially benefit from visual cues. This suggests that there is important cross-modal integration between olfactory and visual modalities [9,12,13]. An ongoing, 4-arm, randomized trial is assessing the best modalities of training to improve olfactory training results [14]. The use of a web application is a promising method for improving olfactory training because it allows for visual stimulation, visual tutorials, the provision of encouragements, and results monitoring. Web applications have been shown to be useful during the SARS-CoV-2 pandemic for triaging patients and assessing trends of the outbreak at a large scale [15-17].

Our patients experienced persistent anosmia for 2 to 12 months, and the rapid recovery that was observed regardless of the anteriority of the anosmia suggests that olfactory improvement was a direct effect of the training. Postinfectious olfactory dysfunction that is not caused by SARS-CoV-2 is associated with moderate rates of spontaneous recovery. Hendriks [18] reported that spontaneous recovery occurs in 35% of patients over a period of approximately 12 months. In a retrospective series of 262 subjects with a mean follow-up time of 14 months, Reden et al [19] reported a 32% improvement in olfactory function, which was assessed with the objective “Sniffin’ Sticks” test, and an increase of at least 6 points in threshold discrimination identification scores. Hummel et al [7] reported a short-term recovery rate of 6% to 8% within 4 months and used the same olfactory tests and definitions for improvement as those of Reden et al [19]. More recently, Havervall et al [20] reported that the incidence rate of olfactory dysfunction after mild SARS-CoV-2 infection among seropositive health care workers was 14.6%, 10.8%, and 9% at 2, 4, and 8 months after infection, respectively, meaning that the spontaneous recovery rate is low [20].

Spontaneous recovery after persistent olfactory dysfunction in patients with SARS-CoV-2 infection is not well described. Vaira et al [4] reported a mean score of 1 on a 10-point analogue subjective olfactory scale (the same one we used) between 30 and 60 days for 138 patients who did not undergo olfactory training, and 20% of patients exhibited olfactory improvement. Our data suggest that improvement can be achieved tardily after 2 months of training. In our study, olfactory training and visual stimulation assisted by a dedicated web application were associated with 73.3% (165/225) of patients whose olfactory function improved by 2 points or more after at least 28 days olfactory training and a mean improvement of 4.4 points [4]. In another study, Lechien et al [21] reported that 15.3% of patients with anosmia and 4.7% of patients with hyposmia did not objectively recover olfaction after 60 days and 6 months, respectively. The comparison of our study with other studies is, however, limited because different olfactory tests and scale evaluations were used [21].

Our study had several limits. There were many excluded patients. Selection bias may exist because we believe that patients who do not feel improvement will more readily stop undergoing training. This could be due to confusion about how benefits are statistically better if patients follow the training regimen for more than 28 days. The mean baseline olfactory function of users who were registered on the web application but underwent less than 7 days of olfactory training or did not record their last olfactory function assessment on web application diary (n=2824; olfactory function score: mean 2.23) was higher than that of the studied population (n=548; olfactory function score: mean 1.9; Student test $P<.001$). The distribution of the olfactory dysfunction severity among patients suggests that patients with more severe olfactory dysfunction from the whole population were retained in the analysis. These data suggest that the results of olfactory training could be better in the whole population than those in the studied population.

There was no control group in this study; therefore, it remains unclear whether the incidence of spontaneous recovery distorted the results. The scale that was used to measure olfactory dysfunction and changes was subjective, as it was a self-assessment analogue scale; scores were self-reported and data about olfactory assessment were not confirmed by physicians and objective tests. However, the possibility of conducting olfactory training at home increased the number of recruited patients and resulted in higher levels of olfactory function recovery compared to those of spontaneous improvement.

Olfactory training and visual stimulation assisted by a dedicated web application was associated with significant olfactive improvement in persistent olfactory dysfunction following SARS-CoV-2 infection, especially after 28 days of olfactory training.

XSL•FO
RenderX

Acknowledgments

We thank the users for their participation in this study as well as Magali Balavoine, MSc, from Weprom in Angers, France, and the anosmie.org patients' association. No one received compensation for their contributions. The sponsor (Weprom) designed and conducted the study; collected, managed, analyzed, and interpreted the data; prepared, reviewed, or approved the manuscript; and made the decision to submit the manuscript for publication.

Authors' Contributions

FD has full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. FD contributed to the concept and design of this study. All authors contributed to the acquisition, analysis, or interpretation of data. All authors contributed to the drafting of the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content. FD and ALS conducted the statistical analysis. FD, Weprom, and Kelindi provided administrative, technical, or material support. FD, SM, and HG supervised this study.

Conflicts of Interest

FD and FL are founders of Kelindi. The other authors have no conflicts of interest.

References

1. Barón-Sánchez J, Santiago C, Martín GGS, Arca R, Fernández R. Smell and taste disorders in Spanish patients with mild COVID-19. Neurologia 2020;35(9):633-638 [FREE Full text] [doi: 10.1016/j.nrl.2020.07.006] [Medline: 32900532]
2. Lee Y, Min P, Lee S, Kim SW. Prevalence and duration of acute loss of smell or taste in COVID-19 patients. J Korean Med Sci 2020 May 11;35(18):e174 [FREE Full text] [doi: 10.3346/jkms.2020.35.e174] [Medline: 32383370]
3. Spinato G, Fabbris C, Polesel J, Cazzador D, Borsetto D, Hopkins C, et al. Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection. JAMA 2020 May 26;323(20):2089-2090 [FREE Full text] [doi: 10.1001/jama.2020.6771] [Medline: 32320008]
4. Vaira LA, Hopkins C, Petrocelli M, Lechien JR, Chiesa-Estomba CM, Salzano G, et al. Smell and taste recovery in coronavirus disease 2019 patients: a 60-day objective and prospective study. J Laryngol Otol 2020 Aug;134(8):703-709 [FREE Full text] [doi: 10.1017/S0022215120001826] [Medline: 32782030]
5. Lechien JR, Chiesa-Estomba CM, Place S, Van Laethem Y, Cabaraux P, Mat Q, COVID-19 Task Force of YO-IFOS. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. J Intern Med 2020 Sep;288(3):335-344 [FREE Full text] [doi: 10.1111/joim.13089] [Medline: 32352202]
6. Santos REA, da Silva MG, Silva MCBDM, Barbosa DAM, Gomes ALDV, Galindo LCM, et al. Onset and duration of symptoms of loss of smell/taste in patients with COVID-19: A systematic review. Am J Otolaryngol 2021;42(2):102889 [FREE Full text] [doi: 10.1016/j.amjoto.2020.102889] [Medline: 33445036]
7. Whitcroft KL, Hummel T. Olfactory dysfunction in COVID-19: Diagnosis and management. JAMA 2020 Jun 23;323(24):2512-2514. [doi: 10.1001/jama.2020.8391] [Medline: 32432682]
8. Damm M, Pikart LK, Reimann H, Burkert S, Göktas Ö, Haxel B, et al. Olfactory training is helpful in postinfectious olfactory loss: a randomized, controlled, multicenter study. Laryngoscope 2014 Apr;124(4):826-831. [doi: 10.1002/lary.24340] [Medline: 23929687]
9. Gottfried JA, Dolan RJ. The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. Neuron 2003 Jul 17;39(2):375-386 [FREE Full text] [doi: 10.1016/s0896-6273(03)00392-1] [Medline: 12873392]
10. Konstantinidis I, Tsakiropoulou E, Constantinidis J. Long term effects of olfactory training in patients with post-infectious olfactory loss. Rhinology 2016 Jun;54(2):170-175. [doi: 10.4193/Rhin15.264] [Medline: 27017331]
11. Hummel T, Rissom K, Reden J, Hähner A, Weidenbecher M, Hüttenbrink KB. Effects of olfactory training in patients with olfactory loss. Laryngoscope 2009 Mar;119(3):496-499. [doi: 10.1002/lary.20101] [Medline: 19235739]
12. Höchenberger R, Busch NA, Ohla K. Nonlinear response speedup in bimodal visual-olfactory object identification. Front Psychol 2015 Sep 30;6:1477 [FREE Full text] [doi: 10.3389/fpsyg.2015.01477] [Medline: 26483730]
13. Amsellem S, Höchenberger R, Ohla K. Visual-olfactory interactions: Bimodal facilitation and impact on the subjective experience. Chem Senses 2018 May 23;43(5):329-339. [doi: 10.1093/chemse/bjy018] [Medline: 29528380]
14. Visual-olfactory training in participants with covid-19 resultant loss of smell. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT04710394 [accessed 2021-05-19]
15. Denis F, Fontanet A, Le Douarin YM, Le Goff F, Jeanneau S, Lescure FX. A self-assessment web-based app to assess trends of the COVID-19 pandemic in France: Observational study. J Med Internet Res 2021 Mar 12;23(3):e26182 [FREE Full text] [doi: 10.2196/26182] [Medline: 33709945]
16. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure FX, et al. Implementation of a self-triage web application for suspected COVID-19 and its impact on emergency call centers: Observational study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]

XSL•FO
RenderX

17. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological observations on the association between anosmia and COVID-19 infection: Analysis of data from a self-assessment web application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]
18. Hendriks APJ. Olfactory dysfunction. Rhinology 1988 Dec;26(4):229-251. [Medline: 3070710]
19. Reden J, Mueller A, Mueller C, Konstantinidis I, Frasnelli J, Landis BN, et al. Recovery of olfactory function following closed head injury or infections of the upper respiratory tract. Arch Otolaryngol Head Neck Surg 2006 Mar;132(3):265-269. [doi: 10.1001/archotol.132.3.265] [Medline: 16549746]
20. Havervall S, Rosell A, Phillipson M, Mangsbo SM, Nilsson P, Hober S, et al. Symptoms and functional impairment assessed 8 months after mild COVID-19 among health care workers. JAMA 2021 May 18;325(19):2015-2016. [doi: 10.1001/jama.2021.5612] [Medline: 33825846]
21. Lechien JR, Chiesa-Estomba CM, Beckers E, Mustin V, Ducarme M, Journe F, et al. Prevalence and 6-month recovery of olfactory dysfunction: a multicentre study of 1363 COVID-19 patients. J Intern Med. Epub ahead of print 2021 Jan 05. [doi: 10.1111/joim.13209] [Medline: 33403772]

Abbreviations

SAS: Statistical Analysis System

Edited by G Eysenbach; submitted 13.04.21; peer-reviewed by F Benezit, S Sopka; comments to author 06.05.21; revised version received 12.05.21; accepted 17.05.21; published 27.05.21.

Please cite as:
Denis F, Septans AL, Periers L, Maillard JM, Legoff F, Gurden H, Moriniere S
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
URL: https://www.jmir.org/2021/5/e29583
doi:10.2196/29583
PMID:34003765

©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), 27.05.2021. 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 Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score

Geza Halasz[1*], MD; Michela Sperti[2*], PhD; Matteo Villani[3], MD; Umberto Michelucci[4], PhD; Piergiuseppe Agostoni[5], MD; Andrea Biagi[1], MD; Luca Rossi[1], MD; Andrea Botti[6], MD; Chiara Mari[6], MD; Marco Maccarini[7], MSc; Filippo Pura[7], MSc; Loris Roveda[7], MSc; Alessia Nardecchia[8], PhD; Emanuele Mottola[9], BA; Massimo Nolli[3], MD; Elisabetta Salvioni[5], MD; Massimo Mapelli[5], MD; Marco Agostino Deriu[2], PhD; Dario Piga[7], PhD; Massimo Piepoli[1], MD, PhD

[1]Department of Cardiology, Guglielmo Da Saliceto Hospital, Piacenza, Italy

[2]PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy

[3]Anesthesiology and ICU Department, Guglielmo da Saliceto Hospital, Piacenza, Italy

[4]TOELT LLC - AI Research and Development, Dubendorf, Switzerland

[5]Department of Clinical Sciences and Community Health, Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico, Milano, Italy

[6]Department of Clinical and Experimental Medicine, University of Parma, Parma, Italy

[7]Dalle Molle Institute for Artificial Intelligence, Università della Svizzera italiana/Scuola universitaria professionale della Svizzera italiana, Lugano, Switzerland

[8]Istituto Istruzione Superiore, Casalpusterlengo, Italy

[9]7HC SRL, Rome, Italy

[*]these authors contributed equally

Corresponding Author:
Geza Halasz, MD
Department of Cardiology
Guglielmo Da Saliceto Hospital
Via Taverna 49,
Piacenza, 29121
Italy
Phone: 39 3517489495
Email: geza.halasz@gmail.com

Abstract

Background: Several models have been developed to predict mortality in patients with COVID-19 pneumonia, but only a few have demonstrated enough discriminatory capacity. Machine learning algorithms represent a novel approach for the data-driven prediction of clinical outcomes with advantages over statistical modeling.

Objective: We aimed to develop a machine learning–based score—the Piacenza score—for 30-day mortality prediction in patients with COVID-19 pneumonia.

Methods: The study comprised 852 patients with COVID-19 pneumonia, admitted to the Guglielmo da Saliceto Hospital in Italy from February to November 2020. Patients' medical history, demographics, and clinical data were collected using an electronic health record. The overall patient data set was randomly split into derivation and test cohorts. The score was obtained through the naïve Bayes classifier and externally validated on 86 patients admitted to Centro Cardiologico Monzino (Italy) in February 2020. Using a forward-search algorithm, 6 features were identified: age, mean corpuscular hemoglobin concentration, PaO_2/FiO_2 ratio, temperature, previous stroke, and gender. The Brier index was used to evaluate the ability of the machine learning model to stratify and predict the observed outcomes. A user-friendly website was designed and developed to enable fast and easy use of the tool by physicians. Regarding the customization properties of the Piacenza score, we added a tailored version of the algorithm to the website, which enables an optimized computation of the mortality risk score for a patient when some of the variables used by the Piacenza score are not available. In this case, the naïve Bayes classifier is retrained over the same derivation cohort but using a different set of patient characteristics. We also compared the Piacenza score with the 4C score and with a naïve Bayes algorithm with 14 features chosen a priori.

XSL•FO
RenderX

Results: The Piacenza score exhibited an area under the receiver operating characteristic curve (AUC) of 0.78 (95% CI 0.74-0.84, Brier score=0.19) in the internal validation cohort and 0.79 (95% CI 0.68-0.89, Brier score=0.16) in the external validation cohort, showing a comparable accuracy with respect to the 4C score and to the naïve Bayes model with a priori chosen features; this achieved an AUC of 0.78 (95% CI 0.73-0.83, Brier score=0.26) and 0.80 (95% CI 0.75-0.86, Brier score=0.17), respectively.

Conclusions: Our findings demonstrated that a customizable machine learning–based score with a purely data-driven selection of features is feasible and effective for the prediction of mortality among patients with COVID-19 pneumonia.

(J Med Internet Res 2021;23(5):e29058) doi:10.2196/29058

KEYWORDS

artificial intelligence; prognostic score; COVID-19; pneumonia; mortality; prediction; machine learning; modeling

Introduction

Despite measureless efforts to limit the spread of COVID-19, over 100 million people have been confirmed positive for SARS-CoV-2 infection and more than 2 million people have died from the virus worldwide, as of February 10, 2021 [1]. While these numbers are rapidly increasing day by day, hospitals have been receiving requests beyond capacity and face extreme challenges concerning a sharp increase in the demand for medical resources as well as a shortage of hospital beds and critical care equipment for the timely treatment of ill patients. Additionally, the clinical spectrum of SARS-CoV-2 infections ranges from asymptomatic status to severe viral pneumonia with respiratory failure and even death, making reliable and successful patient triaging challenging [2].

Data from epidemiological studies suggest that severe illness occurs in approximately 20% of patients and that older age, coexisting medical conditions, and cardiovascular risk factors are associated with worse prognosis [3,4]. In this scenario, identification of the key patient variables driving COVID-19 prognosis is of paramount importance to assist physicians in the early prediction of the pathology trajectory and to improve patient outcomes.

To date, several prognostic models combining clinical and laboratory parameters have been proposed, but they included mainly patients from the first wave of the COVID-19 pandemic. This may cause a risk of bias, making these models unsuitable for clinical decision in daily practice [5,6].

The increasing use of electronic health record (EHR) systems has increased the availability of a large amount of data suitable for machine learning analysis. The latter has already proven its potential to support clinical decisions in many medical fields, including the COVID-19 pandemic [7,8]. Therefore, the aim of this study was to develop and validate a new scoring technique—the Piacenza score—to predict the prognosis of COVID-19 pneumonia, based on a machine learning technique with a purely data-driven selection of prognostic features collected at hospital admission.

We hypothesized that a machine learning score based on data-driven selection of features, which is different from inference statistics, could capture nonlinear relationships among clinical features without human-biased intervention and predict mortality for individual patients more accurately than the currently available risk scores.

Methods

Population and Collected Data

The study was conducted at Guglielmo da Saliceto Hospital, which serves a population of about 300,000 people in the area of Piacenza, Emilia Romagna, in northern Italy. This region has the second highest number of COVID-19 deaths in the country (6219 as of December 7, 2020).

This study retrospectively analyzed the EHRs of a cohort of 852 patients diagnosed with COVID-19 pneumonia according to the World Health Organization interim guidance and admitted to the hospital from February to November 2020. COVID-19 infection was diagnosed by a positive result on a reverse transcriptase–polymerase chain reaction (RT-PCR) assay of a specimen collected on a nasopharyngeal swab. Pregnant women, children (<18 years), and patients with a negative RT-PCR assay were excluded from the study as well as patients presenting with shock and coma.

Data collected in the EHR included patients' demographic information, comorbidities, triage vitals, and laboratory tests and outcomes (including length of stay, discharge, readmission, and mortality). Routine blood examinations at admission comprised complete blood count, coagulation profile, and serum biochemical tests (including renal and liver function, creatine kinase, lactate dehydrogenase, electrolytes, and C-reactive protein). A total of 62 patient characteristics were considered in the score design and development. The study protocol was approved by the local committee on human research.

Criteria for Discharge and Outcome

The criteria for discharge were at the discretion of the caregiver physician. In most cases, the criteria encompassed absence of fever for at least 3 days, as well as substantial clinical improvement including clinical remission of symptoms and 2 throat-swab samples negative for SARS-CoV-2 RNA obtained at least 24 hours apart. The primary outcome was 30-day in-hospital mortality.

Piacenza Score Design

The Piacenza score is a machine learning–based COVID-19 mortality risk predictor. It was implemented using a naïve Bayes approach, which is a probabilistic classifier describing the dependence from the outcome of each variable characterizing the patient, taken separately from the others. The naïve Bayes algorithm was chosen due to the following advantages: (1) it provides a probability of the final outcome, which thus

XSL•FO
RenderX

represents the mortality risk; (2) it can handle both categorical and continuous features; and (3) it can handle missing values, thus providing a mortality risk even when all variable inputs for a patient are not available. Moreover, it proved a successful approach in predicting clinical outcomes in several medical scenarios [9,10]. Other key advantages of using a naïve Bayes classifier are its easy implementation, computational efficiency, optimal scaling performance, and the fact that it achieves good results even in small data sets. Furthermore, it is not influenced by irrelevant features or outliers.

Its major limitation stands in the assumption at the core of the method: features independence. Even if this assumption is almost never satisfied, the classifier proved to reach reasonable results in many scenarios, especially in text classification. Another drawback of naïve Bayes is that if a categorical feature presents a value in the test data set, which was not observed in the training data set, then the model will be unable to make a prediction. Nevertheless, this issue can be solved with various smoothing techniques. Patients missing some features can be easily handled. In fact, only the features' probability distributions need to be computed in training a naïve Bayes classifier. Thus, no imputation was performed and all patients were included in the training phase, since not all missing data were considered for every feature. Furthermore, when applying the trained model to make inferences, the final user can insert missing data, still obtaining a reliable result.

Derivation and Test Cohorts

The EHRs of 852 patients were randomly split into derivation (70%) and test (30%) cohorts. The derivation cohort was first used to select, among the considered 62 patient features, the most significant ones, and then to train the naïve Bayes classifier using only the best predictors, while the predictive ability of the estimated model was assessed on the test cohort.

Piacenza Score Development, Optimization, and Identification of Variable Importance

The Piacenza score has been developed and tailored to (1) minimize the number of clinical variables to be ingested and (2) to maximize the overall prediction performance (ie, in terms of maximization of the area under the receiver operating characteristic curve [AUC]) and patient stratification ability. The most significant patient features were identified through the so-called forward-search approach [11].

The forward-search approach is a purely data-driven dimensionality reduction technique that is able to identify, given a large set of input features, the minimum combination of those features, which maximizes the performance metrics associated with a machine learning algorithm. The forward-search approach was employed here to reduce the number of patient variables from 62 to the 6 most relevant ones used to train the naïve Bayes classifier.

Piacenza Score Evaluation and Metrics

The test cohort was used to assess the performance of the Piacenza score. In order to increase the statistical significance of the results, bootstrapping was used to randomly generate 100 test sets from the original test cohort. Moreover, an external validation cohort has been considered to further validate the Piacenza score performance. The external validation cohort consisted of data from 86 patients with COVID-19 enrolled at Centro Cardiologico Monzino Hospital (Milan, Italy).

The performance of Piacenza score was evaluated in terms of discrimination and calibration capabilities. The discrimination ability was determined by computing the receiver operating characteristic (ROC) curve on the test cohort and the associated AUC, together with its 95% CI. As additional metrics, the negative predictive value (NPV), the positive predictive value (PPV), the accuracy, the sensitivity, the specificity, and the F1 and F2 scores were computed. These metrics were calculated for a threshold value obtained by maximizing the F2 score. The calibration ability was derived by the so-called calibration plots, which compare observed and predicted outcomes with associated uncertainties. The Brier index was used to evaluate the ability of machine learning to stratify and predict observed outcomes. The Brier index is defined as the mean-squared difference between the observed and predicted outcomes and ranges from 0 to 1, with 0 representing the best calibration.

Finally, the variable relative importance was quantified for the identified 6 most relevant patient features. The relative importance is a comparative measure of the patient feature's weight in determining the Piacenza risk score.

Usability, Flexibility, and Customization

The Piacenza score was specifically designed to be an easy, fast, versatile, fair, open, and user-friendly tool. To reach this goal, a web-based calculator of the score, via a website, was released [12]. This calculator can be used by clinicians to estimate a hospitalized patient's risk of 30-day mortality.

We added a tailored version of the algorithm to the website, which enables an optimized computation of the mortality risk score for a patient even when some variables used by the Piacenza score are not available. In this case, the naïve Bayes classifier is retrained over the same derivation cohort but using a different set of patient characteristics. Moreover, a second naïve Bayes model has been presented as a possible example of the Piacenza score's customization and flexibility. The above-mentioned model has been trained with the following 14 variables, chosen a priori by the physician for their association with mortality in COVID-19 pneumonia: age, gender, diabetes, length of symptoms before hospital admission, systolic blood pressure, respiratory rate, PaO_2/FiO_2 ratio, platelets and eosinophils count, neutrophil-to-lymphocyte ratio, C-reactive protein, direct bilirubin, creatinine, and lactate dehydrogenase. Finally, we compared the performance of the Piacenza score with the above-mentioned "clinical" naïve Bayes classifier to show the flexibility of the method, which can be easily retrained with another subset of predictors.

Website Design and Development

The website has been developed in Python (Python Software Foundation), using the Flask framework, and Hosting is managed through Docker.

The site consists of three main pages: *Home*, *Custom Analysis*, and *Multiple Analysis*. The *Home* and the *Custom Analysis*

XSL•FO
RenderX

pages require submitting a form that is dynamically composed in the backend through a Python dictionary variable. This allows us to easily change the form without changing the HTML code. The current dictionary contains the following fields characterizing the features: name, type (continuous or binary), measurement unit, information for the user, value, and mandatory flag.

On the *Home* page, once a form is submitted, the backend receives and sends the parsed data to the previously trained naïve Bayes classifier, which computes the mortality risk that is visualized on the website, typically in less than 1 second. On the *Custom Analysis* page, once a form is submitted and parsed, a naïve Bayes classifier is trained using only the specified features. Since the overall training process may be time-consuming as it also performs feature selection, the final results are automatically sent to the email address specified by the user after completing the training. The *Multiple Analysis* page allows users to compute mortality risks for many patients, without the need to manually fill in a form for every single patient. Clinicians are requested to submit a CSV (comma-separated values) file containing the values of the 6 features characterizing the Piacenza score. An example of the structure of a CSV file is provided on the website.

Comparison With Conventional Risk Models

To further assess the performance of the Piacenza score, we compared it with the 4C mortality score, which considers the following predictors: age, gender, number of comorbidities, respiratory rate, peripheral oxygen saturation (sO_2), level of consciousness (Glasgow coma scale), urea level, and C-reactive protein. The same test cohort used to test the Piacenza score was employed.

Statistical Analysis

Categorical variables were reported as count (%) and continuous variables as mean (SD). A two-sided *P* value <.05 was considered statistically significant. We used the Fisher exact test to assess differences between binary variables and the Welch two-sample *t* test to assess differences between continuous variables. The overall implementation of all codes for the machine learning score and analysis tools was performed in the Python 3.7.4 environment. The Python libraries employed were *pandas* (for data set management), *NumPy* (for numerical computations), and *sklearn* (for data set preprocessing; eg, data set splitting). The naïve Bayes classifier at the core of the Piacenza score was manually implemented (without any additional machine learning framework used) since an existing algorithm for naïve Bayes classification dealing both with continuous and categorical variables as well as missing data was not available in the *sklearn* library. The forward-search algorithm for feature selection was also manually implemented.

Results

Patient Characteristics and Events

A total of 852 patients with SARS-CoV-2 pneumonia were hospitalized during the study period, of which 242 (28%) were admitted to the intensive care unit (ICU). The mean age of the patients was 70 (SD 14) years, and 599 (70%) were male. Comorbidities were present in 602 patients (71%): mainly arterial hypertension (n=499, 59%), dyslipidemia (n=205, 24%), and diabetes (n=157, 18%). The mean time between onset of symptoms and hospital admission was 6.5 (SD 3.9) days. Fever (n=776, 91%), dyspnea (n=543, 64%), and cough (n=400, 47%) were the most common symptoms at admission. A total of 293 patients (34%) died within 30 days after hospital admission. The median time from hospital admission to discharge or death was 9 days. A comparison of clinical characteristics between survivors and nonsurvivors showed that the latter were older (*P*<.001) and had a higher prevalence of hypertension and cerebrovascular disease (*P*<.001); longer symptom duration (*P*<.001); higher respiratory rate (*P*<.001); and lower SpO_2 (*P*<.001), PaO_2/FiO_2 ratio (*P*<.001), and systolic blood pressure at admission (*P*=.02) (Table 1).

XSL•FO
RenderX

Table 1. Study population characteristics and a comparison of survivors and nonsurvivors.

Characteristic	All patients (N=852)	Patients discharged alive (n=559)	Deceased patients (n=293)	*P* value[a]
Gender (male), n (%)	599 (70)	386 (69)	213 (73)	.30
Age (years), mean (SD)	70 (14)	65 (14)	78 (10)	*.001*
Comorbidities, n (%)	602 (71)	364 (65)	238 (81)	*.001*
Hypertension	499 (59)	294 (53)	205 (70)	*.001*
Atrial fibrillation	109 (13)	58 (10)	51 (17)	*.005*
Chronic obstructive pulmonary disease	130 (15)	76 (14)	54 (18)	.07
Dyslipidemia	205 (24)	132 (24)	73 (25)	.67
Chronic kidney disease	75 (9)	42 (8)	33 (11)	.07
Diabetes	157 (18)	90 (16)	67 (23)	*.02*
Cancer	65 (8)	38 (7)	27 (9)	.22
Stroke	28 (3)	9 (2)	19 (6)	*.001*
Peripheral artery disease	19 (2)	10 (2)	9 (3)	.23
Coronary artery disease	96 (11)	58 (10)	38 (13)	.26
Symptoms				
Time from symptom onset to admission, mean (SD)	6.54 (3.94)	6.71 (3.79)	6.27 (4.16)	*.001*
Fever, n (%)	776 (91)	513(92)	263(90)	.32
Dyspnea, n (%)	543 (64)	317(57)	225(77)	*.002*
Cough, n (%)	400 (47)	280 (50)	120 (41)	.18
Fatigue, n (%)	174 (20)	118 (21)	56 (19)	.32
Diarrhea, n (%)	77 (9)	66 (12)	11(4)	.05
Syncope, n (%)	43 (5)	36 (6.5)	7 (2)	.18
Baseline clinical findings, mean (SD)				
PaO_2/FiO_2 ratio	225.93 (96.34)	270.54 (83.82)	196.54 (92.70)	*.001*
pH	7.45 (0.07)	7.46 (0.07)	7.45 (0.07)	.35
PaO_2	60.16 (18.58)	59.68 (15.94)	60.56 (20.54)	.71
$PaCO_2$	35.75 (10.37)	35.36 (8.52)	36.05 (11.58)	.62
HCO_3	25.43 (6.78)	26.22 (9.12)	24.81 (3.97)	.23

[a]*P* value refers to either the Student *t* test or the chi-square test. Italicized values are significant.

Major laboratory markers were tracked upon admission. Specifically, lactate dehydrogenase, creatine kinase, cholinesterase, creatinine, and glycemia were significantly higher in nonsurvivors than survivors (*P*<.001). Nonsurvivors had a significantly lower lymphocyte and eosinophil percentage and red blood cell count as well as lower hemoglobin, mean corpuscular hemoglobin concentration (MCHC), and hematocrit values (*P*<.001). Furthermore, nonsurvivors showed significantly higher levels of inflammatory biomarkers such as neutrophil count, C-reactive protein, and neutrophil-to-lymphocyte ratio (*P*<.001). Other differences in laboratory findings among the two groups are summarized in Table 2.

XSL•FO
RenderX

Table 2. Laboratory findings upon admission for the overall study sample and a comparison of survivors and nonsurvivors.

Laboratory parameter	All patients (N=852), mean (SD)	Patients discharged alive (n=559), mean (SD)	Deceased patients (n=293), mean (SD)	*P* value[a]
Glucose (mg/dl)	145 (66)	137 (59)	159 (76)	*.001*
Urea (mg/dl)	57 (40)	47 (24)	76 (54)	*.001*
Creatinine (mg/dl)	1.24 (0.90)	1.06 (0.54)	1.59 (1.27)	*.001*
Sodium (mEq/l)	137 (8)	137 (8)	137 (7)	.24
Potassium (mEq/l)	4.17 (0.55)	4.14 (0.49)	4.24 (0.65)	*.04*
Chloride (mEq/l)	99.26 (7.21)	98.84 (7.19)	100.05 (7.17)	*.02*
Total bilirubin (mg/dl)	0.75 (0.48)	0.72 (0.35)	0.82 (0.66)	*.02*
Direct bilirubin (mg/dl)	0.22 (0.60)	0.21 (0.69)	0.25 (0.37)	.31
AST[b] (U/L)	61 (84)	53 (37)	79 (136)	*.004*
ALT[c] (U/L)	48 (70)	47 (44)	48 (103)	.90
LDH[d] (U/L)	430 (220)	391 (160)	509 (292)	*.001*
Creatine kinase (U/L)	300 (637)	231 (387)	429 (932)	*.001*
Amylase (U/L)	73 (48)	69 (37)	80 (63)	*.01*
Lipase (U/L)	47 (72)	43 (46)	56 (105)	.06
Serum cholinesterase (U/L)	6275 (1858)	6674 (1763)	5576 (1812)	*.001*
WBC[e] × 10^3/µl	8.12 (4.68)	7.86 (4.72)	8.63 (4.56)	*.02*
RBC[f] × 10^6/µl	4.69 (0.72)	4.79 (0.68)	4.51 (0.77)	*.001*
Hemoglobin (g/dl)	13.59 (1.91)	13.83 (1.72)	13.14 (2.16)	*.001*
Hematocrit (%)	41.84 (5.70)	42.37 (5.34)	40.83 (6.22)	*.001*
MCV[g] (fl)	89.74 (6.66)	89.18 (5.62)	90.80 (8.19)	*.003*
MCH[h] (pg)	29.13 (2.38)	29.05 (2.12)	29.28 (2.80)	.23
MCHC[i] (g/dl)	32.43 (1.36)	32.56 (1.15)	32.17 (1.66)	*.001*
Platelets × 10^3/µl	217.75 (117.90)	221.08 (127.10)	211.41 (97.72)	.22
RDW[j] (%)	13.65 (1.65)	13.27 (0.27)	14.29 (1.99)	*.001*
Neutrophils (%)	77.45 (11.57)	75.81 (11.75)	80.56 (10.55)	*.001*
Lymphocytes (%)	15.17 (9.20)	16.48 (9.45)	12.67 (8.15)	*.001*
Monocytes (%)	6.89 (4.30)	7.16 (4.01)	6.36 (4.76)	*.02*
Eosinophils (%)	0.32 (0.91)	0.38 (1.05)	0.20 (0.54)	*.001*
Lymphocytes × 10^3/µl	1.09 (0.99)	1.15 (0.94)	0.98 (1.09)	*.03*
Monocytes × 10^3/µl	0.51 (0.41)	0.52 (0.35)	0.51 (0.51)	.77
Eosinophils × 10^3/µl	0.02 (0.07)	0.03 (0.08)	0.02 (0.05)	*.04*
Neutrophils × 10^3/µl	6.41 (3.72)	6.05 (3.41)	7.11 (4.15)	*.001*
PT[k] (seconds)	15.84 (8.38)	15.07 (5.83)	17.03 (11.11)	*.02*
Prothrombin activity (%)	68.40 (15.96)	69.86 (14.38)	66.27 (17.82)	*.009*
INR[l]	1.40 (0.76)	1.34 (0.65)	1.51 (0.93)	*.01*
PTT[m] (seconds)	31.70 (5.74)	31.32 (4.48)	32.29 (7.22)	.08
PTT ratio	1.02 (0.19)	1.00 (0.14)	1.04 (0.25)	.06
C-reactive protein (mg/dl)	11.19 (8.55)	9.85 (7.88)	13.74 (9.17)	*.001*

XSL•FO
RenderX

Laboratory parameter	All patients (N=852), mean (SD)	Patients discharged alive (n=559), mean (SD)	Deceased patients (n=293), mean (SD)	*P* value[a]
NLR[n]	7.99 (6.74)	6.78 (5.04)	10.27 (8.68)	*.001*

[a]*P* value refers to either the Student *t* test or the chi-square test. Italicized values are significant.

[b]AST: aspartate aminotransferase.

[c]ALT: alanine aminotransferase.

[d]LDH: lactate dehydrogenase.

[e]WBC: white blood cell count.

[f]RBC: red blood cell count.

[g]MCV: mean corpuscular volume.

[h]MCH: mean corpuscular hemoglobin.

[i]MCHC: mean corpuscular hemoglobin concentration.

[j]RDW: red cell distribution width.

[k]PT: prothrombin time.

[l]INR: international normalized ratio.

[m]PTT: partial thromboplastin time.

[n]NLR: neutrophil-to-lymphocyte ratio.

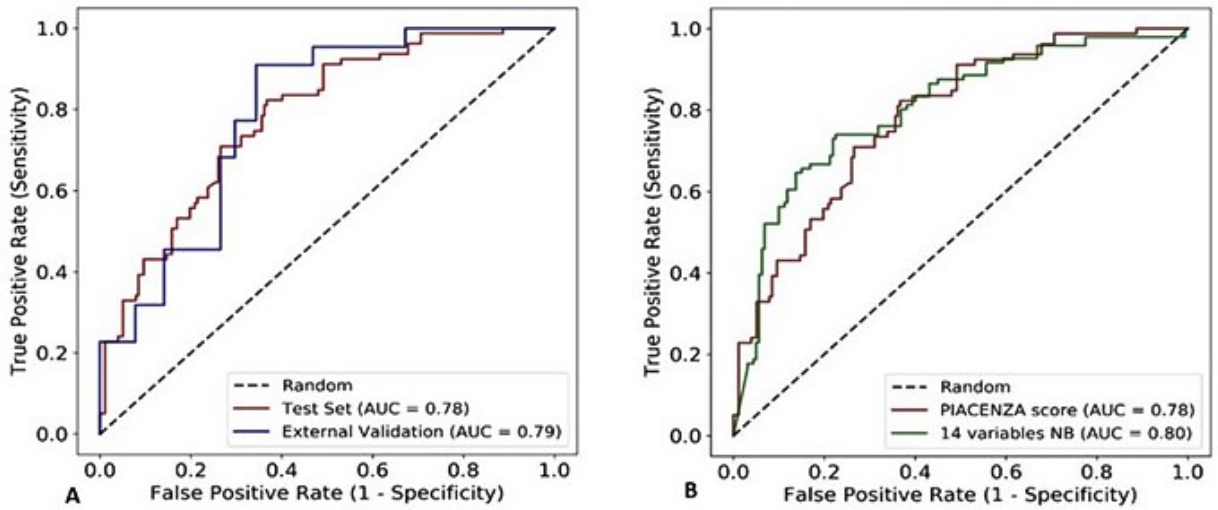
Significant Predictors and the Piacenza Score

Using the forward-search algorithm, the following 6 most important predictors at hospital admission were identified and used to compute the Piacenza score: age, MCHC, PaO_2/FiO_2 ratio, temperature, previous cerebrovascular stroke, and gender.

The median of the ROC curve over 100 test cohorts (generated through bootstrapping) is reported in Figure 1. The corresponding median of the AUC is equal to 0.78 (95% CI 0.74-0.84) with a sensitivity of 94% and specificity of 37%. The NPV of the Piacenza score was 93% with a PPV of 40% (Table 3).

The calibration plot of the Piacenza score over the range of risk showed a Brier score of 0.19. The risk deciles are grouped into three levels: low risk (first to fifth deciles), intermediate risk (sixth to eighth deciles), and high risk (ninth and tenth deciles). A gradual and progressive increase in absolute event rates was observed across risk classes for all the Piacenza scores (death: 14% [18/125] in low-risk deciles vs 36% [27/75] in intermediate-risk deciles vs 66% [33/50] in high-risk deciles).

Figure 1. (A) Receiver operating characteristic (ROC) curves obtained by evaluating the Piacenza score (red curve) on the test cohort and on the external validation cohort. (B) ROC curves obtained by evaluating the Piacenza score (red curve) and the naïve Bayes (NB) model trained with 14 manually chosen features (green curve). AUC: area under the ROC curve.



XSL•FO
RenderX

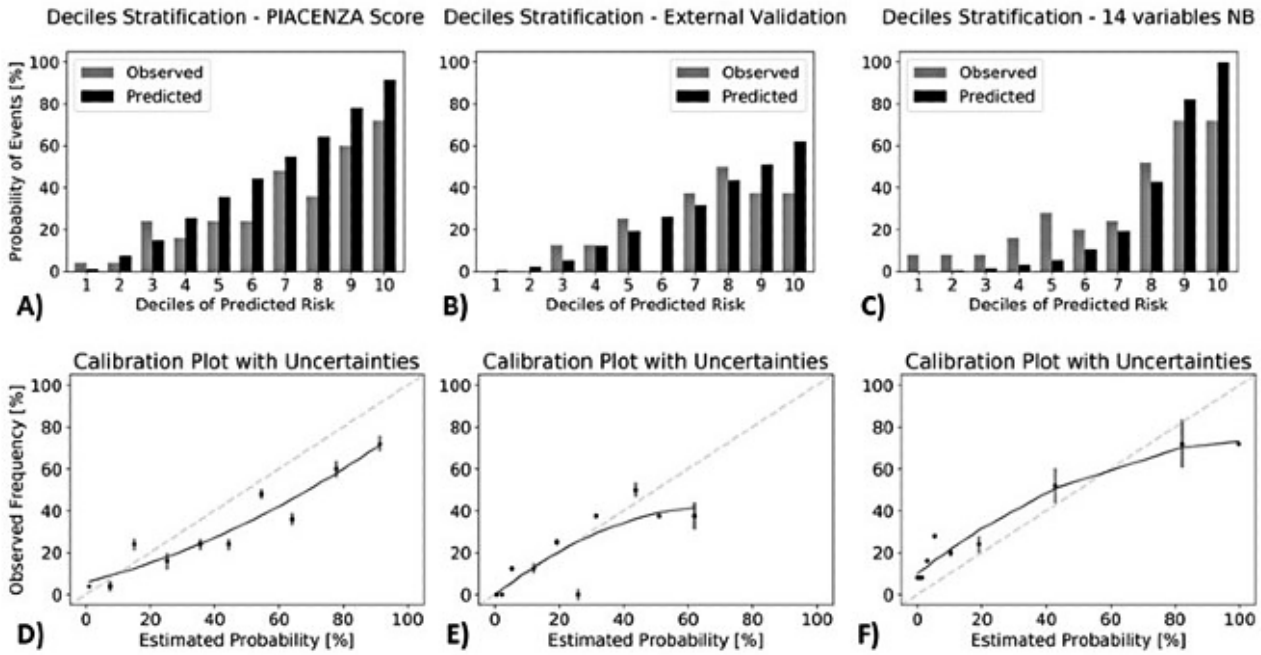
Table 3. Negative predictive value (NPV), positive predictive value (PPV; or precision), accuracy, sensitivity (or recall), specificity, F1 score, and F2 score for all scores. These metrics have been calculated for a specific threshold value on the final risk score probability chosen by maximizing the F2 score, the reason being that F2 privileges a high recall and therefore a broader confidence for correctly identifying patients at risk.

Scores	Threshold	NPV	PPV	Accuracy	Sensitivity	Specificity	F1 score	F2 score
Piacenza score	0.16	0.93	0.40	0.55	0.94	0.37	0.56	0.74
Piacenza score–external validation	0.16	0.97	0.37	0.57	0.95	0.44	0.53	0.72
Naïve Bayes model trained with 14 manually chosen features	0.04	0.88	0.54	0.67	0.88	0.55	0.67	0.78
4C mortality score	0.12	0.98	0.39	0.53	0.99	0.34	0.56	0.76

From the computed calibration plot, we can observe that the mortality risk is underestimated only in the first few deciles, while in the higher deciles the risk is slightly overestimated (Figure 2A-D).

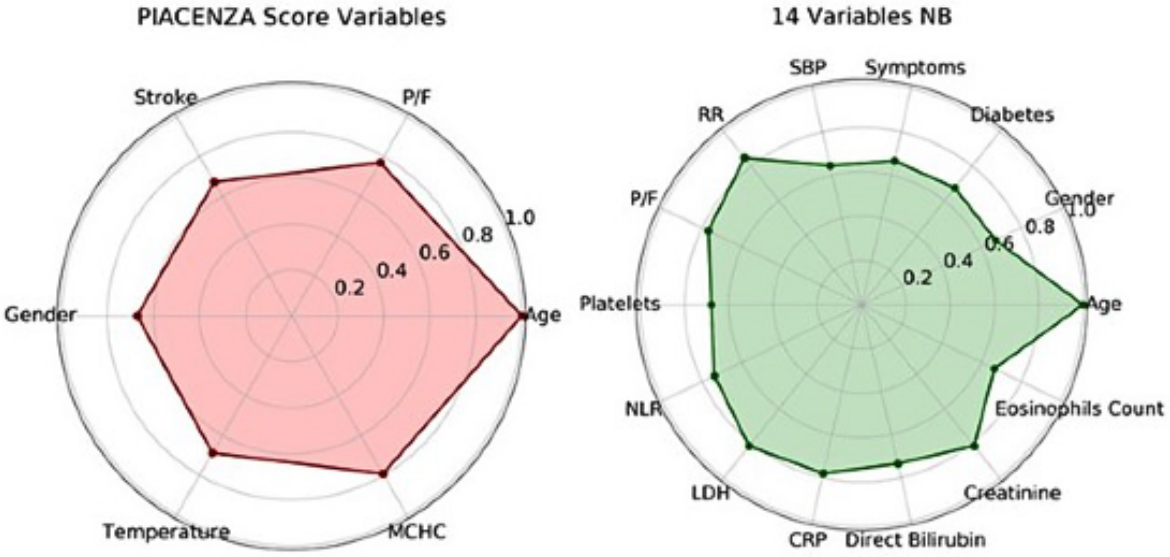
Regarding the relative importance of each features independent from the others, age was the most important feature to predict death followed by MCHC, PaO_2/FiO_2 ratio, previous cerebrovascular stroke, gender, and temperature (Figure 3).

Figure 2. Risk of observed death according to deciles of event probability based on the Piacenza score (A), the Piacenza score on the external validation data set (B), and the naïve Bayes (NB) model trained with 14 manually chosen features (C). For every single case, the corresponding calibration plots with standard deviations calculated over the deciles are also shown below each respective graph (D, E, and F).



XSL•FO
RenderX

Figure 3. Radar plot for the 6 Piacenza score predictors of death and for the 14 manually chosen features, showing their relative importance. Feature importance is scaled with respect to the most important feature. NB: naïve Bayes, MCHC: mean corpuscular hemoglobin concentration, CRP: C-reactive protein, LDH: lactate dehydrogenase, NLR: neutrophil-to-lymphocyte ratio, P/F: PaO2/FiO2, RR: respiratory rate, SBP: systolic blood pressure.



External Validation

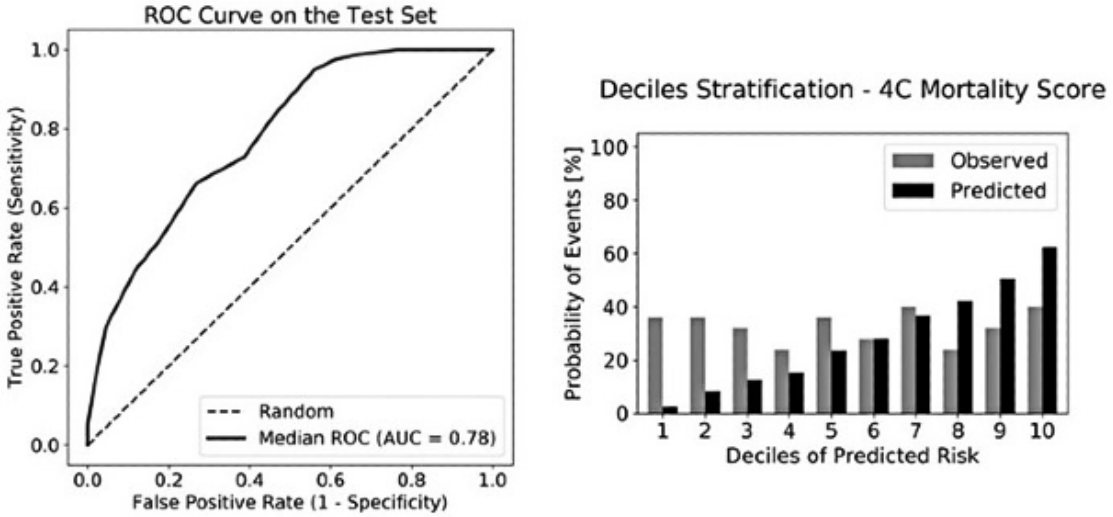
The corresponding median of the AUC in the external validation cohort was 0.79 (95% CI 0.68-0.89) with a Brier score of 0.16 (Figure 1A), a sensitivity of 95%, and a specificity of 44% (Table 3).

The calibration plot is reported in Figure 2B and showed again a gradual and progressive increase in absolute event rates across risk classes (death: 10% [4/40] in low-risk deciles vs 29% [7/24] in intermediate-risk deciles vs 38% [6/16] in high-risk deciles).

Comparison With the 4C Mortality Score and the Naïve Bayes Model Using Manually Chosen Features

The median of the AUC was 0.78 (95% CI 0.73-0.83) with a sensitivity of 99% and specificity of 34% for the 4C score when evaluated on the test cohort. The corresponding Brier score was equal to 0.26 (Figure 4). The naïve Bayes model with 14 features chosen manually based on clinician experience achieved an AUC of 0.80 (95% CI 0.75-0.86) with a sensitivity of 88%, a specificity of 55%, and a Brier score of 0.17 (Figure 1B). The detailed performance metrics of both scores are reported in Table 3. The relative importance of the selected 14 features of the naïve Bayes model is shown on the radar plot in Figure 3.

Figure 4. Performance of the 4C mortality score (both in terms of discrimination and calibration abilities) calculated on the test cohort. ROC: receiver operating characteristic, AUC: area under the ROC curve.



The observed mortality increased gradually and progressively for the naïve Bayes model with manually chosen features—death: 14% (17/125) in low-risk deciles vs 32% (14/75) in intermediate-risk deciles vs 72% (36/50) in high-risk deciles. This was not observed for the 4C score—death: 33% (41/125) in low-risk deciles vs 31% (23/75) in intermediate-risk deciles vs 36% (18/50) in high-risk deciles. Both scores achieved a satisfactory patient stratification only in the last three deciles whereas the 4C mortality score overestimated the prediction in the high-risk deciles and underestimated it in the low-risk ones (Figures 2C-4).

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we developed and validated a machine learning–based risk score—the Piacenza score—to predict mortality risk among hospitalized patients with COVID-19 pneumonia. This score is based on only 6 variables that are readily available at hospital admission.

Satisfactory performance, measured in terms of AUCs in both the testing and external validation cohorts, was achieved with excellent patient stratification. More specifically, the Piacenza score showed a higher sensitivity with a lower specificity. Likewise, it underestimated the mortality risk in the first three risk deciles; slight overestimation occurred in the other deciles. This behavior is acceptable and preferred in an acute setting since the score has been designed as a screening predictive tool capable of correctly identifying patients at low risk from those at high risk of mortality.

In crowded hospitals, and with shortages of medical resources, this simple model can help to quickly prioritize patients: if the patient's estimated risk is low, the clinician may choose to monitor the patient, whereas a high-risk estimate might support aggressive treatment or admission to the ICU. Data from China, Europe, and the United States reported a hospitalization rate of 20% to 31%, an ICU admission rates from 17% to 35%, and an in-hospital mortality rate between 15% and 40% [13]. In our study, the in-hospital 30-day mortality rate was 34% with lower survival rates for older patients with pre-existing comorbidities and with clinical signs and symptoms suggesting respiratory failure at hospital admission. In line with previous findings, we found that the most common laboratory abnormalities among patients who died were related to the inflammatory process, renal and liver damage, and procoagulation status [14,15].

In the presence of a large number of patients requiring intensive care and threatening to overwhelm health care systems around the world, several models to predict survival and guide clinical decisions in COVID-19 pneumonia were developed [16]. However, many of these models have been found to have a high risk of bias, which could reflect their development based on a small study population with high risk of overfitting and poor generalization properties to new cohorts, and without clear details of model derivation and testing [6].

The recent spread of artificial intelligence has brought novel ways to combat current global pandemics by collecting and analyzing large amounts of data, identifying trends, stratifying patients on the basis of risk, and proposing solutions at the population level instead of at the single individual level [17,18].

Comparison With Other Risk Stratification Scores

During the COVID-19 pandemic, machine learning approaches have been used to predict the outbreak, to diagnose the disease, to analyze chest x-ray and CT (computed tomography) scan images, and more recently to predict mortality or progression risk to severe respiratory failure [19,20].

Yuan and colleagues [21] developed a simple prognostic risk score based on a logistic regression classifier that included 3 laboratory markers: lactate dehydrogenase, high-sensitivity C-reactive protein, and lymphocyte percentage. This score was developed from a cohort of 1479 patients and externally validated in 2 independent cohorts, reaching an accuracy of 95% in predicting the risk of mortality. However, the model comprised only Chinese patients during the early stages of the outbreak and, more importantly, it seems to have a significant selection bias as it did not include patients with mild and moderate disease at admission [21].

The 4C mortality score, developed and validated by the International Severe Acute Respiratory and Emerging Infections Consortium, based on 8 clinical and laboratory variables, achieved an AUC of 0.78 in predicting mortality. It is easy to use and has a pragmatic design. In fact, to calculate the score, no external tool or complex mathematical equation is required, and results can be immediately retrieved at the bedside [5]. However, due to the rapidly evolving characteristics of the virus and its impact on the population, the score should be continuously updated. For example, the 4C score did not include patients from the second wave of the pandemic. At the same time, if a broad range of individuals are included, the score may become unsuitable for more specific clinical scenarios, such as patients affected by severe pneumonia.

The performance of our model is comparable with the 4C mortality score applied to the test cohort used in this paper. However, we remark that the 4C mortality score was derived based on a population of 35,000 patients, while the naïve model providing the Piacenza score was trained using information coming only from 852 patients. This is indicative of the high representativeness of the training cohort considered in our study. Furthermore, although there is a similar discriminative power between the 4C score and the Piacenza score, the latter score showed better performance in stratifying patients according to their mortality risk, which is of paramount importance in selecting the appropriate treatment and for resource allocations. We also externally tested our score, achieving good performance and confirming that our data-driven model is robust despite its reliance on variables deemed relevant in this context without actually knowing their semantics.

Characteristics of the Piacenza Score

The Piacenza score contains parameters reflecting patient demographics, comorbidity, and physiology at hospital admission. It shares some characteristics with the 4C score such as age, gender, comorbidities, and PaO_2/FiO_2 but also includes unexplored features like temperature and MCHC deriving from a substantially different selection of variables. Unlike traditional scores based on logistic regression analysis mixed with a knowledge-driven approach where a score is assigned by an expert to each of the limited number of selected variables, the proposed predictive model is purely data driven and is not affected by a clinically oriented, potentially biased choice of variables [22].

The Piacenza score is highly customizable and can be adapted as more information becomes available on disease progression and the impact of interventions like vaccines and new pharmacological treatments. In fact, the naïve Bayes algorithm, during its learning phase, generates a summary of the data set

XSL•FO
RenderX

where each variable is associated with the outcome in terms of a probabilistic dependence. This summary describes the distribution of the current data set and can be quickly and easily updated when a new observation is available, adapting itself to changes within the population. Likewise, if new data are available, they can be used to train a new version of the Piacenza score and study the possible fingerprints of COVID-19 variants.

The Piacenza score is thus highly flexible; if the some of the required variables are missing, the model can be retrained and the physician can still receive a customized result (associated with the best possible accuracy with respect to the variables provided). The retraining process can take up to 10 hours, depending on the number of features inserted. However, depending on future requests, codes can be easily optimized and run on more powerful hardware.

An example of a personalized model different from the Piacenza score is the naïve Bayes model trained with 14 manually chosen features, which showed a predictive power comparable to that of the Piacenza score. Other models differ in performance; however, as demonstrated, the variables age and PaO_2/FiO_2 ratio have the biggest contribution to the predictive power of the model. Therefore, starting with age and the PaO_2/FiO_2 ratio and adding more variables will lead to predictive performances similar to that of the Piacenza score, which represents the best combination for stratifying patients and predicting mortality.

Finally, our score's predictors were not chosen a priori (like, for example, the 4C mortality score) but as the product of a machine learning–based optimization technique, which considers the smallest possible subset of leading predictors associated with the best possible performance.

The Piacenza Score Beyond the COVID-19 Pandemic

The approach proposed in our paper is suitable for risk stratification and mortality assessment of other conditions as well, such as heart failure (HF), which constitutes a growing public health issue. In fact, although machine learning has made significant contributions to health care in just a few years, little evidence exists on the role of machine learning in predicting mortality in patients with HF and in general with cardiovascular diseases. In this context, several researchers have developed prognostic risk scores for HF such as the Seattle Heart Failure Model and the Meta-Analysis Global Group in Chronic Heart Failure [23,24]. However, these models do not necessarily predict mortality in patients with HF at the individual level and do not present the same flexibility as the Piacenza score. When dealing with cardiovascular diseases, the flexibility of the scores is of crucial importance due to the continuous and rapid changes in therapeutic strategies; this makes the above-mentioned scores less useful or not reliable in clinical practice.

Limitations

This study has room for further improvement, which is left for future work. First, given that the proposed machine learning method is purely data driven, our model may vary if a different data set is used. As more data become available, the model can be refined and performance of the Piacenza score can further increase. To this aim, we are currently looking forward to subsequent large-sample and multicentered studies. Second, the forward-selection algorithm (used to select the Piacenza score predictors and most importantly to personalize the Piacenza score on any other subset of features) may be an expensive option to be considered and may surely be optimized in further versions of the code. Finally, new variables such as d-dimer and troponin, currently not available, but which are known to be associated with a higher mortality risk in cases of COVID-19 pneumonia may be included in future analyses.

Conclusion

In conclusion, we have developed and validated robust machine learning models, which could be used to predict the prognosis of patients with COVID-19. The Piacenza score has several advantages: first, it relies on objective clinical and laboratory measurements not affected by human interpretation; second, it was tested and validated in patients belonging to the second wave of the pandemic; third, it is automatically generated through a combination of variables widely available at hospital admission and can be calculated through a user-friendly web interface; and finally, as opposed to traditional epidemiological predictive models, the Piacenza score has the added advantage of adaptive learning, trend-based recalibration, and flexibility.

Acknowledgments

No sponsor had any role in the study design, data collection, data analysis, data interpretation, or writing of the paper.

Authors' Contributions

GH, DP, MV, and MAD conceived the study. A Biagi, LR, A Botti, CM, AN, MM, and ES collected the data. MS, UM, AN, and DP managed and analyzed the data. MM, FP, LR, and EM developed the website. PA and MN provided clinical expertise. MP supervised the work. All authors interpreted the results. GH, MAD, MV, and DP wrote the manuscript, which was approved by all the authors.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)? China CDC Weekly 2020;2(8):113-122 [FREE Full text] [doi: 10.46234/ccdcw2020.032]
2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/nejmoa2002032]
3. 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. The Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/s0140-6736(20)30566-3]
4. Biagi A, Rossi L, Malagoli A, Zanni A, Sticozzi C, Comastri G, et al. Clinical and epidemiological characteristics of 320 deceased patients with COVID-19 in an Italian Province: A retrospective observational study. J Med Virol 2020 Nov 29;92(11):2718-2724 [FREE Full text] [doi: 10.1002/jmv.26147] [Medline: 32515500]
5. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Correction: Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020 Nov 13;371:m4334 [FREE Full text] [doi: 10.1136/bmj.m4334] [Medline: 33187971]
6. Wynants L, Van Calster B, Collins G, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
7. 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]
8. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr 2020 Jul;14(4):337-339 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.012] [Medline: 32305024]
9. Wolfson J, Bandyopadhyay S, Elidrisi M, Vazquez-Benitez G, Vock DM, Musgrove D, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. Stat Med 2015 Sep 20;34(21):2941-2957 [FREE Full text] [doi: 10.1002/sim.6526] [Medline: 25980520]
10. Kanwar MK, Gomberg-Maitland M, Hoeper M, Pausch C, Pittrow D, Strange G, et al. Risk stratification in pulmonary arterial hypertension using Bayesian analysis. Eur Respir J 2020 Aug 04;56(2):2000008 [FREE Full text] [doi: 10.1183/13993003.00008-2020] [Medline: 32366491]
11. Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. Pattern Recognition Letters 1994 Nov;15(11):1119-1125. [doi: 10.1016/0167-8655(94)90127-9]
12. Halasz G, Sperti M, Villani M, Michelucci U, Agostoni P, Biagi A, et al. The Piacenza Score. URL: https://covid.7hc.tech [accessed 2021-05-18]
13. Wiersinga W, Rhodes A, Cheng A, Peacock S, Prescott H. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
14. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. Life Sci 2020 Aug 01;254:117788 [FREE Full text] [doi: 10.1016/j.lfs.2020.117788] [Medline: 32475810]
15. Lippi G, Plebani M. Laboratory abnormalities in patients with COVID-2019 infection. Clin Chem Lab Med 2020 Jun 25;58(7):1131-1134. [doi: 10.1515/cclm-2020-0198] [Medline: 32119647]
16. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical Predictive Models for COVID-19: Systematic Study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]
17. Fu Z, Hong S, Zhang R, Du S. Artificial-Intelligence-Enhanced Mobile System for Cardiovascular Health Management. Sensors (Basel) 2021 Jan 24;21(3):773 [FREE Full text] [doi: 10.3390/s21030773] [Medline: 33498892]
18. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. The Lancet 2021 Jan;397(10270):199-207. [doi: 10.1016/s0140-6736(20)32519-8]
19. Ferrari D, Milic J, Tonelli R, Ghinelli F, Meschiari M, Volpi S, et al. Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia-Challenges, strengths, and opportunities in a global health emergency. PLoS One 2020 Nov 12;15(11):e0239172 [FREE Full text] [doi: 10.1371/journal.pone.0239172] [Medline: 33180787]
20. Yadaw AS, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health 2020 Oct;2(10):e516-e525. [doi: 10.1016/s2589-7500(20)30217-x]
21. Yuan Y, Sun C, Tang X, Cheng C, Mombaerts L, Wang M, et al. Development and Validation of a Prognostic Risk Score System for COVID-19 Inpatients: A Multi-Center Retrospective Study in China. Engineering (Beijing) 2020 Nov 28 [FREE Full text] [doi: 10.1016/j.eng.2020.10.013] [Medline: 33282444]
22. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Acad Emerg Med 2016 Mar 13;23(3):269-278 [FREE Full text] [doi: 10.1111/acem.12876] [Medline: 26679719]

XSL•FO
RenderX

23. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 2006 Mar 21;113(11):1424-1433. [doi: 10.1161/CIRCULATIONAHA.105.584102] [Medline: 16534009]
24. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. Meta-Analysis Global Group in Chronic (MAGGIC) Heart Failure Risk Score: Validation of a Simple Tool for the Prediction of Morbidity and Mortality in Heart Failure With Preserved Ejection Fraction. J Am Heart Assoc 2018 Oct 16;7(20):e009594 [FREE Full text] [doi: 10.1161/JAHA.118.009594] [Medline: 30371285]

Abbreviations

AUC: area under the receiver operating characteristic curve
CSV: comma-separated values
CT: computed tomography
EHR: electronic health record
HF: heart failure
ICU: intensive care unit
MCHC: mean corpuscular hemoglobin concentration
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
RT-PCR: reverse transcriptase–polymerase chain reaction

Edited by G Eysenbach; submitted 29.03.21; peer-reviewed by D Lazzeroni, X Cheng, M Elbattah, P Zhao, L Li; comments to author 18.04.21; revised version received 15.05.21; accepted 16.05.21; published 31.05.21.

Please cite as:
Halasz G, Sperti M, Villani M, Michelucci U, Agostoni P, Biagi A, Rossi L, Botti A, Mari C, Maccarini M, Pura F, Roveda L, Nardecchia A, Mottola E, Nolli M, Salvioni E, Mapelli M, Deriu MA, Piga D, Piepoli M
A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score
J Med Internet Res 2021;23(5):e29058
URL: https://www.jmir.org/2021/5/e29058
doi:10.2196/29058
PMID:33999838

©Geza Halasz, Michela Sperti, Matteo Villani, Umberto Michelucci, Piergiuseppe Agostoni, Andrea Biagi, Luca Rossi, Andrea Botti, Chiara Mari, Marco Maccarini, Filippo Pura, Loris Roveda, Alessia Nardecchia, Emanuele Mottola, Massimo Nolli, Elisabetta Salvioni, Massimo Mapelli, Marco Agostino Deriu, Dario Piga, Massimo Piepoli. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2021. 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-Stratified Infection Probabilities Combined With a Quarantine-Modified Model for COVID-19 Needs Assessments: Model Development Study

Vena Pearl Bongolan[1], PhD; Jose Marie Antonio Minoza[1], BSc; Romulo de Castro[2], PhD; Jesus Emmanuel Sevilleja[3], MD, MPhil

[1]Department of Computer Science, University of the Philippines Diliman, Quezon City, Philippines

[2]Center for Informatics, University of San Agustin, Iloilo, Philippines

[3]National Center for Mental Health, Mandaluyong, Philippines

Corresponding Author:
Vena Pearl Bongolan, PhD
Department of Computer Science
University of the Philippines Diliman
UP AECH Bldg
Velasquez St
Quezon City, 1101
Philippines
Phone: 63 915 877 2298
Email: bongolan@up.edu.ph

Abstract

Background: Classic compartmental models such as the susceptible-exposed-infectious-removed (SEIR) model all have the weakness of assuming a homogenous population, where everyone has an equal chance of getting infected and dying. Since it was identified in Hubei, China, in December 2019, COVID-19 has rapidly spread around the world and been declared a pandemic. Based on data from Hubei, infection and death distributions vary with age. To control the spread of the disease, various preventive and control measures such as community quarantine and social distancing have been widely used.

Objective: Our aim is to develop a model where age is a factor, considering the study area's age stratification. Additionally, we want to account for the effects of quarantine on the SEIR model.

Methods: We use the age-stratified COVID-19 infection and death distributions from Hubei, China (more than 44,672 infections as of February 11, 2020) as an estimate or proxy for a study area's infection and mortality probabilities for each age group. We then apply these probabilities to the actual age-stratified population of Quezon City, Philippines, to predict infectious individuals and deaths at peak. Testing with different countries shows the predicted number of infectious individuals skewing with the country's median age and age stratification, as expected. We added a Q parameter to the SEIR model to include the effects of quarantine (Q-SEIR).

Results: The projections from the age-stratified probabilities give much lower predicted incidences of infection than the Q-SEIR model. As expected, quarantine tends to delay the peaks for both the exposed and infectious groups, and to "flatten" the curve or lower the predicted values for each compartment. These two estimates were used as a range to inform the local government's planning and response to the COVID-19 threat.

Conclusions: Age stratification combined with a quarantine-modified model has good qualitative agreement with observations on infections and death rates. That younger populations will have lower death rates due to COVID-19 is a fair expectation for a disease where most fatalities are among older adults.

(J Med Internet Res 2021;23(5):e19544) doi:10.2196/19544

KEYWORDS

COVID-19; epidemic modeling; age stratification theory; infection probability; SEIR; mathematical modelling

XSL•FO
RenderX

Introduction

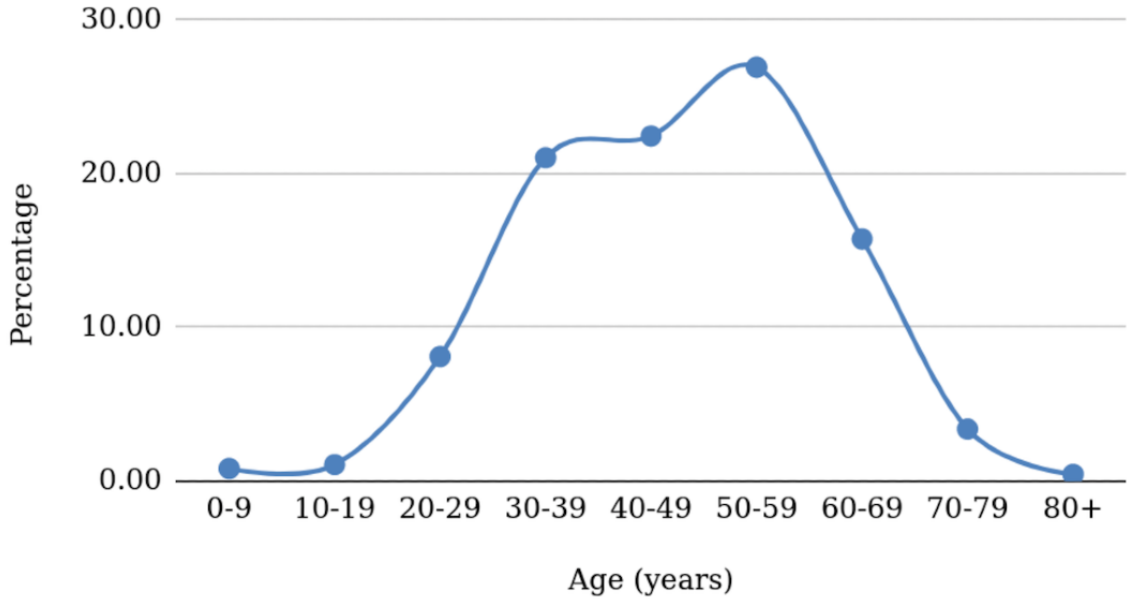
The initial impression that came out of Wuhan, China, in late 2019 and early 2020 was that COVID-19 most affects older adult males with pre-existing conditions. Classic compartmental models like the susceptible-exposed-infectious-removed (SEIR) model all assume a homogenous population, and that everyone has equal chances of getting infected. The SEIR model is initialized by "dividing" the population into four compartments; people "progress" through being susceptible, to getting exposed, to being infectious, to getting removed, either via recovery with permanent immunity or death. Permanent immunity is a common assumption when modelling viral infections, and it is assumed here. In future, the model may be modified for temporary immunity, as soon as we get reliable data on reinfection rates. A scan of the preprints from various modeling efforts during the first quarters of 2020 gave high estimates for the peaks of the exposed and infectious groups (40% of the population, by some estimates). Even our quarantine-modified model suffered from this, and this inspired us to use age-stratified infection probabilities, which gave us a lower bound for estimates.

Methods

Estimates by Age Stratification

This calculates the "mathematical expectation" of future infections per age group, by multiplying an age group's infection probability by the population in that age group. Initially, we used the data of patients with COVID-19 in Hubei [1], stratified by ages. As data came in, we repeated the calculations with updated Quezon City data. We treat the percentages of incidence in each age group as a proxy or estimate for the corresponding probabilities of infection for people in the corresponding age group. The true probabilities are and continue to be unknown, but the scatter of the data from Hubei is consistent with the virus affecting older adults with pre-existing conditions more than other groups (Figure 1).

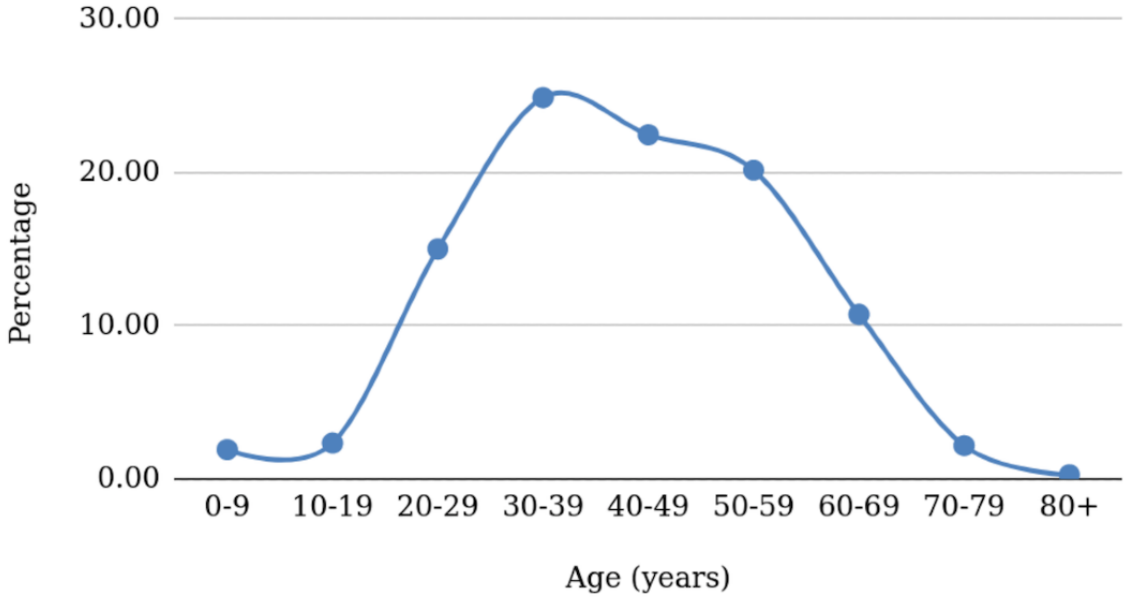
Figure 1. Projected age-stratified percentages of cases for China.



Next, we took the proxy probabilities from Hubei, and applied them to the actual age stratification of the Philippines. Due to its young population, this resulted in a "skewed to the right" distribution (Figure 2) compared to the Hubei distribution, and the true distribution for the study area will be revealed as actual cases are reported.

XSL•FO
RenderX

Figure 2. Projected age-stratified percentages of cases for the Philippines.



The Philippines has a median age of 25.7 years [2] compared to China's 38.4 years [2]. This means half of the population is aged <25.7 years, so more than half of the population will be in the "safer" age groups, with lower probabilities of getting infected, and significantly greater chances of survival if they should contract the disease. Those who do get infected account for only 10.2% of cases (see Table 1, sum of the percentages of those aged 0-29 years old). We expect this to be true for other countries with low median ages.

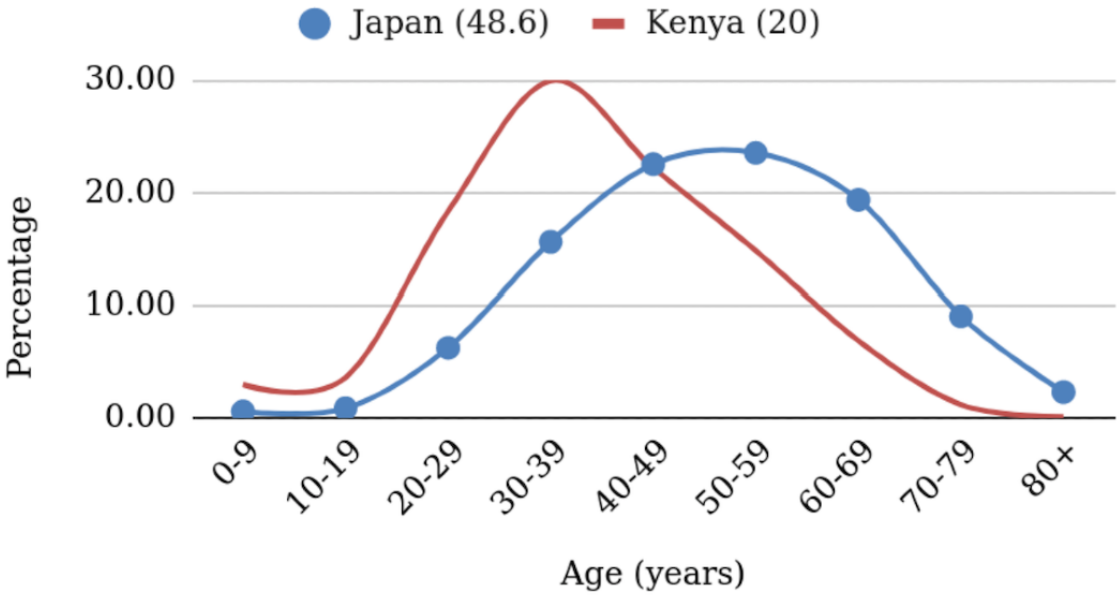
Using the United Nations World Population Prospects 2019 data [3], we did a similar experiment with Japan (median age 48.6 years) and Kenya (median age 20 years) [2]. We see the younger population also skewing right (Figure 3). Martinez [4] did similar calculations, but did not treat the age-stratified incidence as a proxy for infection probability.

Table 1. Quezon City, Philippines, projection of COVID-19 cases and mortality using COVID-19 data from Hubei, China.

Age group (years)	Hubei, China		Quezon City, Philippines		
	COVID-19 cases, %	Case fatality rate, %	Projected COVID-19 cases, n	Projected COVID-19 case distribution, %	Projected mortality, n
0-9	0.9	0	5164.60	1.60	0.00
10-19	1.2	0.2	7345.98	2.28	14.69
20-29	8.1	0.2	53,444.14	16.57	106.89
30-39	17	0.2	87,156.90	27.02	174.31
40-49	19.2	0.4	75,013.55	23.25	300.05
50-59	22.4	1.3	61,753.09	19.14	802.79
60-69	19.2	3.6	27,420.07	8.50	987.12
70-79	8.8	8.0	4588.82	1.42	367.11
≥80	3.2	14.8	699.83	0.22	103.57
Total	100	28.7	322,586.99	100	2856.54

XSL•FO
RenderX

Figure 3. Projected age-stratified percentages of cases for Japan and Kenya. The average age in each country is given in parentheses.



Estimates Using a Quarantine-Modified SEIR Model

The quarantine-modified SEIR equations are shown below.



The insertion of the *Q(t)* term serves to control the $S \times I$ or susceptible-infectious interactions, hence lowering exposure. *Q(t)* equal to one means no quarantine (ie, no change to the model). A *Q(t)* value of 0.4 means a 60% effective quarantine, while *Q(t)*=0.6 means the quarantine is only 40% effective; therefore, the lower the *Q(t)*, the better. We allowed *Q(t)* to vary day by day (since cases began before the quarantine), and we also estimated the success of the quarantine. Henceforth, we refer to this model as Q-SEIR. Solution was via the Euler method, and time stepping was one day. We used Excel (Microsoft Corp) worksheets for our calculations.

The model was ground-truthed to the estimated number of exposed individuals at the national level on April 1, 2020 (N=7400) [5]. From the nationally reported number of exposed individuals (patients under investigation + persons under monitoring), Quezon City represents almost 10% of cases (~740); the Q-SEIR model predicted 705 cases.

Results

Applying the Hubei infection probabilities on Quezon City with an age distribution as shown in Table 1 (from the 2015 Census, projected to 2020 at a 2% growth rate [6]) gave an estimate of 322,586 infectious individuals (accumulated, which we equate with the Q-SEIR peak), which accounts for less than 10% of the population of Quezon City. Deaths were predicted at 22,390 cases (6.94%), which lies between the World Health Organization (WHO) morbidity estimates of 5.58% [7] and the 12.47% reported in Italy [8]. These estimates were what was available as of April 2020, and will be updated in the discussion.

Discussion

Principal Results

Initial reports (around April 2020) estimated the Philippines' death or mortality rate to be at 4.70% (4.05%-5.43%) [8]. This high estimate may be explained by sampling bias, wherein severe cases may have been overrepresented because of a lack of testing. Those who are infectious but are asymptomatic or exhibit mild symptoms should also be equally represented in the testing guidelines (at the moment, they are not), as well as those who were infectious with no symptoms and have recovered.

We tried the calculations from [4] using the death rate, which was reported at 2.3% for China [1]. This gave a much lower number of around 2857 deaths, for a Quezon City death rate of 0.89%. This figure is surprisingly low compared to the 6.94% projected using the estimated infection probability. The latest estimate of the Philippine COVID-19 death rate is 1.82% [9].

The delay in test reporting (estimate of 5-7 days [10]) factors in the estimation of the initial E-I-R values. In addition, this delay is compounded by the incubation period and, in our opinion, moves the quarantine effect further down from the actual date of implementation (March 15, 2020). We started the Q-SEIR simulation on March 20, 2020, with no quarantine assumed because the steep jump in cases occurred on this date; 60% effective quarantine was set for April 2, 2020.

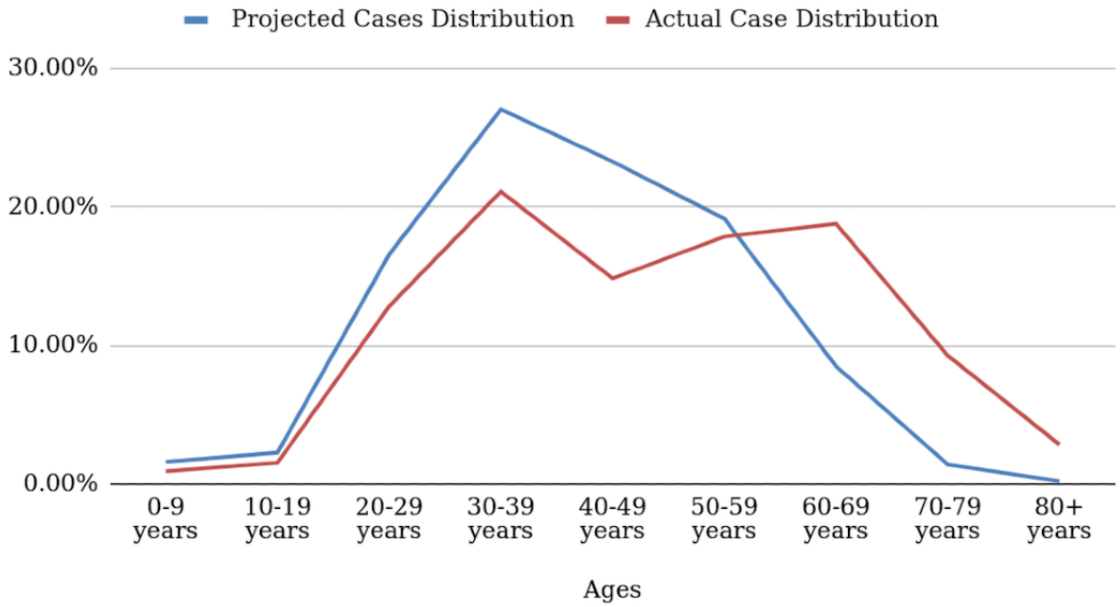
The Q-SEIR model predicted 14.00% of the population will be infectious (I) at the peak. The two methods now give us a low and high estimate for Quezon City: infectious individuals will peak between 9.95% (from age stratification) and 14.00% (from Q-SEIR) of the population, around the third week of May 2020. These figures seemed high compared to the reported incidence during the same period, but were not in any way unique (compared to modeling done in other countries). At that time, the suspicion was that actual cases were undertested/underreported by as much as 90% (ie, only 10% of

XSL•FO
RenderX

cases were being detected). Nevertheless, this range of values serves as a guide for planners in anticipating the need for personal protective equipment, mass testing, hospital beds, and other basic needs.

Figure 4 shows the scatter of cases for Quezon City projected by the Hubei data, and the actual scatter as of May 2020.

Figure 4. Projected versus actual age-stratified case distribution for Quezon City.



Limitations

Like many research groups, we were and continue to be hampered by a lack of reliable data in a usable format. Even now, we refer to multiple data sets including the Philippine Department of Health's DataDrop, WHO, and other data sources.

Comparison With Prior Work

We were one of the first groups to forward the theory of age stratification when it comes to modeling COVID-19 infections. Recently, we came across works by Balabdaoui et al [11], Undurraga et al [12], and the WHO [13]. However, we are uniquely using incidence percentages as proxies for infection probabilities with good results.

Conclusions

In conclusion, age stratification predicted the scatter of cases for Quezon City fairly well. It also predicted later observations of lower cumulative confirmed cases for the same time period (eg, 6587 cases per million compared to Italy's 58,417 cases per million as of March 28, 2021); there is a color-coded world map in [14]. Some African countries have lower median ages than the Philippines, and they have generally lighter colors than Europe and North America. The best prediction is the noticeably lower death or mortality rate of 1.82% compared to Italy's 3.05% and Indonesia's 2.7% (all as of March 29, 2021) [9]. Since COVID-19 disproportionately affects older adults more than younger populations, we expect the Philippines to have a lower mortality rate than countries with older populations (eg, Italy has a median age of 47.3 years and Indonesia has a median age of 29.7 years, compared to 25.7 years in the Philippines). Later work can use actual infection probabilities to include the effect of age stratification in the model.

Acknowledgments

VPB thanks the Alumni Engineers of the University of the Philippines Diliman for their support of the study (UPCoE Covid-19 Response). Additionally, she thanks her coauthors, RdC and JES, who worked pro bono; true heroes! Finally, the authors are grateful to the Engineering Research and Development for Technology (ERDT) Consortium, Republic of the Philippines, for shouldering publication costs.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020[J]. China CDC Weekly 2020 Mar 03;2(8):113-122. [doi: 10.46234/ccdcw2020.032]
2. Central Intelligence Agency. Field Listing - Median age. The World Factbook. URL: https://www.cia.gov/the-world-factbook/field/median-age/ [accessed 2020-04-11]
3. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019. URL: https://population.un.org/wpp/DataQuery/ [accessed 2020-04-11]
4. Martinez R. COVID-19 mortality calculator. Public Tableau. URL: https://public.tableau.com/profile/ramon.martinez#!/vizhome/COVID-19mortalitycalculator/COVID-19mortalitycalc [accessed 2020-04-06]
5. Department of Health. COVID-19 Case Tracker. URL: https://www.doh.gov.ph/covid-19/case-tracker [accessed 2020-04-01]
6. Philippines: Metro Manila. City Population. URL: https://www.citypopulation.de/php/philippines-metromanila-admin.php?adm2id=137404 [accessed 2020-04-06]
7. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-04-06]
8. Oke J, Heneghan C. Global Covid-19 Case Fatality Rates. The Centre for Evidence-Based Medicine. URL: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/ [accessed 2020-04-06]
9. Coronavirus (COVID-19) death rate in countries with confirmed deaths and over 1,000 reported cases, by country. Statista. URL: https://www.statista.com/statistics/1105914/coronavirus-death-rates-worldwide/ [accessed 2020-04-01]
10. COVID-19 test results from RITM out in 5 to 7 days, but not for long, DOH says. CNN Philippines. URL: https://www.cnnphilippines.com/news/2020/3/27/COVID-19-test-results-from-RITM-out-in-5-to-7-days,-but-not-for-long,-DOH-says.html [accessed 2020-04-01]
11. Balabdaoui F, Mohr D. Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland. Sci Rep 2020 Dec 04;10(1):21306 [FREE Full text] [doi: 10.1038/s41598-020-77420-4] [Medline: 33277545]
12. Undurraga E, Chowell G, Mizumoto K. COVID-19 case fatality risk by age and gender in a high testing setting in Latin America: Chile, March-August 2020. Infect Dis Poverty 2021 Feb 03;10(1):11 [FREE Full text] [doi: 10.1186/s40249-020-00785-1] [Medline: 33531085]
13. World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. 2020 Mar 17. URL: https://apps.who.int/iris/handle/10665/331656 [accessed 2021-05-12]
14. Philippines: Coronavirus Pandemic Country Profile. Our World in Data. URL: https://ourworldindata.org/coronavirus/country/philippines [accessed 2020-04-01]

Abbreviations

SEIR: susceptible-exposed-infectious-removed
WHO: World Health Organization

Edited by C Basch; submitted 22.04.20; peer-reviewed by N Mohammad Gholi Mezerji, D Cebo; comments to author 26.10.20; revised version received 31.03.21; accepted 05.04.21; published 31.05.21.

Please cite as:
Bongolan VP, Minoza JMA, de Castro R, Sevilleja JE
Age-Stratified Infection Probabilities Combined With a Quarantine-Modified Model for COVID-19 Needs Assessments: Model Development Study
J Med Internet Res 2021;23(5):e19544
URL: https://www.jmir.org/2021/5/e19544
doi:10.2196/19544
PMID:33900929

©Vena Pearl Bongolan, Jose Marie Antonio Minoza, Romulo de Castro, Jesus Emmanuel Sevilleja. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2021. 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 of Mobile Health Apps and Features to Support Psychosocial Well-being Among Frontline Health Care Workers Involved in the COVID-19 Pandemic Response: Qualitative Study

Sungwon Yoon[1], MPH, PhD; Hendra Goh[1], BSc; Gayathri Devi Nadarajan[2], MBBS; Sharon Sung[1], PhD; Irene Teo[1,3], PhD; Jungup Lee[4], PhD; Marcus E H Ong[1,2], MBBS, MPH, FRCS; Nicholas Graves[1], PhD; Tess Lin Teo[2], MBBS

[1]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[2]Department of Emergency Medicine, Singapore General Hospital, Singapore, Singapore

[3]Lien Centre for Palliative Care, Duke-NUS Medical School, Singapore, Singapore

[4]Department of Social Work, Faculty of Arts and Social Sciences, National University of Singapore, Singapore, Singapore

Corresponding Author:
Sungwon Yoon, MPH, PhD
Health Services and Systems Research
Duke-NUS Medical School
8 College Rd, Singapore 169857
Singapore,
Singapore
Phone: 65 66013198
Email: sungwon.yoon@duke-nus.edu.sg

Abstract

Background: Frontline health care workers are experiencing a myriad of physical and psychosocial challenges amid the COVID-19 pandemic. There is growing recognition that digital technologies have the potential to improve the well-being of frontline workers. However, there has been limited development of wellness interventions using mobile health (mHealth) technology. More importantly, little research has been conducted on how frontline workers perceive mHealth-based support to promote their well-being.

Objective: This study aimed to explore frontline workers' experience of conventional psychological wellness programs and their perceptions of the usefulness of mHealth apps and features for promoting well-being. It also sought to identify factors that could potentially influence uptake and retention of an mHealth-based wellness program.

Methods: We conducted semistructured interviews using purposive sampling with frontline workers involved in the COVID-19 response. Various visual materials, collated from existing mHealth app features, were presented to facilitate discussion. Interviews were audio-recorded and transcribed verbatim. Thematic analysis based on grounded theory was undertaken. Themes were subsequently mapped to key nudge strategies—those commonly used for mHealth development—to assess participants' preferences for particular features and their reasoning.

Results: A total of 42 frontline workers participated in 12 one-on-one interviews or focus group discussions. Frontline workers generally had a limited ability to identify their own psychological problems and liked the *reminders* functionality of the app to track their mood over time. A personalized goal-setting feature (ie, *tailoring*) and in-app *resources* were generally valued, while frequent coaching and messages (ie, *framing*) were seen as a distraction. The majority of participants desired a built-in chat function with a counselor (ie, *guidance*) for reasons of accessibility and protection of privacy. Very few participants appreciated a *gamification* function. Frontline workers commonly reported the need for ongoing social support and desired access to an in-app peer support community (ie, *social influence*). There were, however, concerns regarding potential risks from virtual peer interactions. Intrinsic motivational factors, mHealth app technicality, and tangible rewards were identified as critical for uptake and retention.

Conclusions: Our study highlights the potential of mHealth apps with relevant features to be used as wellness tools by frontline health care workers. Future work should focus on developing a nonintrusive and personalized mHealth app with in-app counseling, peer support to improve well-being, and tangible and extrinsic rewards to foster continued use.

(J Med Internet Res 2021;23(5):e26282) doi:10.2196/26282

XSL•FO
RenderX

KEYWORDS

COVID-19; frontline health care workers; mHealth; well-being; psychosocial

Introduction

As the COVID-19 pandemic continues to claim more lives, health systems globally have been severely strained. The consequence has been mounting pressure on frontline health care workers [1]. Being at the heart of delivering essential health services during the pandemic, frontline workers face an increased burden and risk of developing physical and psychosocial problems [2]. Studies of previous infectious disease crises, such as SARS, H1N1, and Ebola, reported that frontline workers suffered from poor emotional and social functioning, depressive symptoms, and insomnia, all in addition to the distress relating to the risk of being infected and infecting their loved ones [3-5]. Likewise, emerging literature on COVID-19 consistently indicates a myriad of physical, psychological, and social challenges faced by frontline workers, including burnout, posttraumatic stress symptoms, anxiety, and depression [6,7].

Learning from the experience of prior public health emergencies, it has been suggested that support for the mental well-being of frontline workers is a requisite strategy to ensure a healthy and resilient workforce capable of providing a sustained COVID-19 response. Many institutions worldwide have begun to develop wellness intervention plans to support frontline workers during the pandemic [8-10]. These programs have included specialized mental care clinics, employee assistance counseling, and hospital-based interactive activities [11,12]. However, with few exceptions, the majority of the psychosocial support and wellness programs have been based on in-person interventions or one-off support, without much consideration given to sustained monitoring and impact assessment.

Organizations such as NHS (National Health Service) Digital in the United Kingdom and the US Department of Health and Human Services have underlined the importance of using mobile technology to improve access to care [13,14]. There is also a growing recognition that information and communication technologies have the potential to address the many challenges faced by the public and frontline workers during public health emergencies [15]. This technology focus has led to a surge in the development of mobile health (mHealth) apps during the ongoing COVID-19 pandemic [16,17]. According to recent review studies, a total of 29 mHealth apps pertaining to COVID-19 have been developed by 19 countries [18]. Yet, all of them were exclusively designed for contact tracing and symptom management with the primary purpose of containing disease transmission.

Despite increased rhetoric and enthusiasm for wellness support for frontline workers via virtual platforms, there remain limited advances toward developing psychosocial interventions using mHealth technology for this population. More importantly, little research so far has been undertaken on how frontline workers perceive mHealth-based support to meet their psychological needs and challenges during the ongoing COVID-19 situation. In light of the dearth of literature, this study aimed to explore frontline workers' experience of conventional psychosocial wellness programs and their perceptions of the usefulness of mHealth apps and various features for promoting well-being. The study also sought to identify factors that could potentially influence uptake and retention of mHealth-based wellness support programs. Exploring frontline workers' perspectives as end users will inform future efforts to design evidence-based and user-centered mHealth interventions.

In this study, we adopted nudge theory as a way to better understand perceptions of mHealth app features designed to support psychosocial well-being. Nudge theory is a novel concept that affects the process of decision making in individuals through choice architecture [19]. The theory has been widely used to develop mHealth apps in general and has also been applied to interventions promoting mental well-being. A growing body of literature has demonstrated evidence of nudge elements that promote health behavior and optimize user outcomes with varying degrees of effectiveness in mHealth interventions [20]. For example, a study by Wiecek et al highlighted that use of *reminders* is a successful intervention component for mHealth apps to improve medication adherence, as participants are regularly "nudged" to take their medication [21]. Martin et al called for the incorporation of *gamification* in an mHealth app, as it motivates the user to adopt intended behavior through gamified activities [22]. *Social influence* has also been considered an important nudge that increases the effects of intervention outcomes by fostering a feeling of belonging to a community with a common goal [23]. Individualized *tailoring* has been found to be an effective nudge in sustaining user engagement for digital mental health interventions [24-26]. In addition, provision of app-based resources yielded a positive outcome that alleviated chronic stress of working women [27]. Taken together, nudge theory has proved to be useful in effecting behavior change for mHealth interventions. Ahead of the design of an intervention to support the well-being of frontline workers, we sought to explore their views and preferences for app features that underpin these key nudge strategies.

Methods

Setting and Participants

This qualitative study was conducted in Singapore during the peak of the COVID-19 outbreak from April to August 2020. Singapore is a multiethnic city-state in Southeast Asia. Although the country's public health care system consists of a team of highly trained health care professionals and efficient services, the surge in COVID-19 infections placed considerable strains on the health system [28]. Mass outbreaks in foreign worker dormitories in April 2020 spurred the government's decision to impose a 2-month lockdown [29]. The evolving outbreak situation posed unprecedented risks that could compromise the well-being of frontline workers. In response, a multidisciplinary task force of the Singapore General Hospital, the largest public hospital in Singapore, swiftly rolled out a wellness program aimed at mitigating the risk of burnout and adverse stress

XSL•FO
RenderX

reactions among frontline workers. The program included operational and psychological education, provision of a written handout for family members, active support of staff who were involved in high-risk areas, and close communication between peer support leaders and clinical team members.

It is against this background that a study was conducted to explore perceptions of frontline workers (ie, doctors and nurses) about mHealth apps to improve their well-being. The study was introduced to clinical teams at the Singapore General Hospital and Community Care Facilities—large-scale institutional isolation units—at their staff meetings. Participants were identified using official websites, the study team's professional networks, and recommendations from other study participants. In this study, we defined frontline workers as those who were (1) doctors and nurses and (2) in direct contact with suspected and confirmed COVID-19 patients and provided essential health care services. Potential participants were invited to participate by email and were provided with background information. We used a purposive sampling approach to maximize the diversity of experiences and opinions. As the data collection and concurrent analyses progressed, the variation in emergent themes was explored by recruiting subsequent participants for focus groups and interviews based on profession and years of experience to improve our understanding of specific aspects of the studied phenomenon. Prior to interviews, informed consent was sought via email.

Data Collection

A semistructured interview guide was developed based on relevant literature and the study team's expert knowledge [30-32]. Major topics included perceptions of the existing wellness program available for frontline workers, perceived usefulness of mHealth apps for mental well-being, features that might be valuable for improving wellness, and factors affecting adoption of mHealth apps for wellness. In addition, self-reported demographic information, such as age, gender, ethnicity, and years of experience in health care, were collected. During the interview, various visual materials, collated from existing mHealth app features for wellness promotion, were presented to facilitate the discussion. These materials were organized according to key nudge strategies commonly used for mHealth app development [33,34]. They included, but were not limited to, the following strategies: framing, tailoring, reminders, social influence, guidance, social modeling, gamification, and resources [19]. These strategies enabled a structured way to collect participants' perspectives on the usefulness of mHealth features for improving well-being. Due to the constraints of participants' working hours, consented individuals took part in either a one-on-one interview or focus group discussion, subject to availability. For safety reasons, all interviews were conducted virtually over Zoom by two interviewers trained in qualitative research (SY and HG). Reflections and memos were written after each interview to capture insights. Interviews lasted 31 to 65 minutes. This study was approved by the National University of Singapore Institutional Review Board (NUSIRB No: S-20-115)

Data Analysis

All interviews were audio-recorded following consent and were transcribed verbatim. Thematic data analysis was undertaken using grounded theory. A grounded theory approach based on the work of Glaser and Strauss was chosen because it explored experiential aspects of psychosocial programs and how participants perceived the usefulness of key features in a wellness app. The grounded theory approach allowed for emerging constructs and themes through iterations of data collection and analysis [35,36]. Two independent coders (SY and HG) reviewed the interview materials, summarized and extracted meaningful statements, and carried out open coding and axial coding using NVivo 12 (QSR International), a qualitative data analysis software. During open coding, transcripts were analyzed to develop categories of information. This allowed for subthemes to be derived from the data instead of pre-existing ideas. During axial coding, common subthemes were grouped into unifying themes. The iterative process of independent coding and consensus meetings continued until no new emergent themes were identified. The codes were independently applied to all transcripts, and coding discrepancies were resolved by iterative discussions. Lastly, for mHealth features, themes were mapped against the nudge strategies to assess participants' preferences for particular features and their reasoning [34]. For rigor and transparency, we anchored our methodology according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Multimedia Appendix 1) [37].

Results

Characteristics of Participants

A total of 42 frontline workers participated in 12 one-on-one interviews or focus group discussions. Two one-on-one interviews and 10 focus group discussions were conducted, comprising 21 doctors and 21 nurses (for composition of participants in each session, see Table S1 in Multimedia Appendix 2). The recruitment rate was 93% (42/45). A total of 3 individuals out of 45 (7%) declined participation for reasons of lack of time and disinterest. During the peak of the COVID-19 outbreak, health care workers from a variety of clinical departments had been deployed to frontline clinical duties. Participants' clinical home departments ranged across anesthesiology, dentistry, community nursing, and cardiology, among others. Data saturation was reached after the 10th interview, with no new themes emerging from subsequent interviews. We conducted two additional focus groups beyond data saturation to ensure that the point of information redundancy had been achieved. Table 1 shows the characteristics of participants, including 21 (50%) doctors and 21 (50%) nurses out of 42; 76% (32/42) of participants were female and 64% (27/42) were Chinese. The mean age of the sample was 29.6 (SD 3.9) years, with slightly more than half of the participants (24/42, 57%) aged 30 years and below. Participants' experience in the health care sector ranged from 2 to 18 years.

XSL•FO
RenderX

Table 1. Characteristics of participants.

Characteristic	Value (N=42)
Age (years)	
Mean (SD)	29.6 (3.9)
Category, n (%)	
<30	24 (57)
30-50	18 (43)
Ethnicity, n (%)	
Chinese	27 (64)
Malay	7 (17)
Indian	5 (12)
Others	3 (7)
Gender, n (%)	
Female	32 (76)
Male	10 (24)
Profession, n (%)	
Doctor	21 (50)
Nurse	21 (50)
Years of experience in health care, mean (SD)	6.8 (3.6)

Experience and Perceptions of the Conventional Psychosocial Program

Table 2 shows the experience of participants regarding the existing wellness program and their coping strategies. By and large, the wellness program was perceived to be appropriate, with some participants expressing positive thoughts about the educational nature of the session. In particular, junior doctors who had never experienced an outbreak response appreciated the opportunity to learn how to protect themselves and where to find resources in the event of burnout. However, as the program comprised a one-off session delivered to a mixture of different health care teams, some expressed the need for building rapport and relationship to enhance the experience. As one participant noted, the session was felt to be a "random group meeting with strangers" (Participant #19, doctor). A minority of participants were not aware of the program's existence. The inherent absence of team spirit and social connectivity, coupled with limited awareness, seemed to discourage frontline workers from active involvement. Most participants preferred that the program session be organized according to clinical teams.

A recurring theme was that informal help-seeking appeared to be the most common avenue to ameliorate the emotional "fallout" from work for our participants. The majority indicated seeking support from family and friends when there was "emotional exhaustion." This was more apparent for participants who had limited awareness of the formal support program or felt that the current program did not meet their expectations. Importantly, participants commonly noted that they often could not recognize their own symptoms of burnout. The cumulative psychological effects of burnout at times unwittingly resulted in high anger expression and feelings of frustration. When prompted, some participants cited constant fatigue and physical exhaustion amid longer work hours as the main reason that limited their ability to self-identify psycho-emotional signs and symptoms.

XSL•FO
RenderX

Table 2. Perception of in-person wellness program and personal coping practice.

Theme and subthemes	Illustrative quotes
Experience of formal wellness program	
Greater need for building rapport and relationships	"I think the program was okay. But the group was huge, so I think sometimes people might be a bit shy to share feelings in a large group." (Participant #1, doctor)
	"The social workers gave us their contact details after the session, which was good as we can call whenever we need help. But I feel some people might not utilize them as there was no prior rapport built." (Participant #35, nurse)
Limited awareness of wellness program for uptake	"I was not aware of that [wellness program], but I do know that there are such services for the COVID-19 patients." (Participant #11, doctor)
	"I don't think there was a [wellness program], maybe I missed the email. Not too sure about that." (Participant #42, doctor)
Lacking team spirit and social connectivity among colleagues	"The session was slightly awkward as it was just out of the blue. It was like a random group meeting with strangers, I feel." (Participant #19, nurse)
	"It might be better to split the sessions into participants who are from the same team, as some people might be uncomfortable sharing in the presence of people whom they might not know." (Participant #1, doctor)
Personal coping practice	
Informal help-seeking or non–help-seeking as a prevailing practice	"Let's say on that particular day if I am feeling frustrated regarding work, I will simply talk to my family members, and they are very supportive." (Participant #4, doctor)
	"I am sure the hospital has put in place some sort of program to help us psychologically, just that I never actively sought for help." (Participant #17, doctor)
Limited ability to self-identify psychoemotional symptoms	"Sometimes when I am so tired, I just don't know what to feel anymore." (Participant #39, nurse)
	"There was this time when I had a mini rage as everything just built up. Until a senior noticed this and she came over to ask me what happened and helped me to resolve the issues." (Participant #14, doctor)

Perceived Usefulness of Features in Wellness mHealth Apps

Table 3 presents perceived usefulness of various features for an mHealth app. Across seven nudge strategies (ie, reminders, tailoring, guidance, framing, social influence or modeling, gamification, and resources), 14 themes emerged. It was commonly perceived that using *reminders* to nudge individuals to monitor their own psychological well-being would be helpful in improving self-awareness, particularly when tracking was visualized in a graphic form for easy interpretation. As one participant noted, the tracking feature would allow for "reflection on stressors" (Participant #4, doctor) by alerting users of the presence of mood deterioration. However, frequent notifications were seen as an "annoyance." In addition to the tracking feature, participants valued a personalized goal-setting feature (ie, *tailoring*), which could aid them by altering them to diet and sleep patterns. Almost all the participants desired a built-in chat with a counselor (ie, *guidance*), as it would offer greater convenience and enable ready access to professional care compared to the conventional mode of in-person counseling. However, many participants did not appreciate the artificial intelligence–driven chatbot, due to the loss of human interactivity. They expressed hesitance toward this feature, although some recognized that it might be useful for simple tasks such as requests for mental health resources.

To enhance the user experience, the acceptability of a *framing* feature was explored. Given frequent changes in work protocols in light of the COVID-19 situation and its consequential work demands, occasional messages and coaching were favored compared to daily push notifications, which were seen as an "annoyance." Participants felt that the former would offer a sense of human touch without causing intrusion into private life. Under the *social influence* nudge strategy, features such as forum chats and in-app peer support groups were seen as an essential component of psychosocial support. Despite the positive perceptions of the social influence features, participants highlighted that proper safeguarding measures should be in place to ensure appropriate balance between freedom of expression and emotional manipulation, especially by users who post malicious comments. Related to this, some participants had concerns about disclosure of their personal well-being state that might inadvertently affect assessment of their work performance. This concern was more salient among junior doctors.

The *gamification* function, a widely used nudge strategy to improve self-efficacy, was not something participants were keen to use. While a minority saw a gaming strategy as entertainment, most believed that competition via scoring may be demotivating and backfire when individuals fall behind their peers and colleagues. Virtual rewards such as earning badges, intended for triggering competitive natures, were generally received in a negative light. App-based *resources*, such as mindfulness-based exercises and short wellness articles, were highly valued as a handy and useful tool in promoting a healthier lifestyle and mental well-being.

XSL•FO
RenderX

Table 3. Perceived usefulness of features in wellness mobile health apps.

Nudge strategy and categories	Theme	Illustrative quotes
Reminders (ie, timely cues)		
Mood monitoring	Regular in-app tracking of a mood trend perceived as a tool for improving self-awareness of emotions	"I think the feature can help me to record when is the day that I am feeling low or happy." (Participant #2, doctor)
Progress tracking	Regular in-app tracking of a mood trend perceived as a tool for improving self-awareness of emotions	"You can keep track of your progress and monitor your mood. Eventually, you can see a trend which might potentially help you to identify stressors in daily life. I won't mind using it." (Participant #5, doctor)
Push notification for engagement	Frequent notifications being felt as a distraction	"I do not like this idea [periodical receipt of push notifications], as I find it [to be] annoying. But as long as there is this option of turning it off, then it is fine." (Participant #11, doctor)
Tailoring (ie, context-sensing based on input)		
Tailored feedback	Feedback system aids users in making beneficial changes	"I think this feature [feedback] is a good thing. I wouldn't mind using it, as it allows me to understand better where I have done great and in which area I can work to improve my mental well-being." (Participant #3, doctor)
Personalized goal setting	Setting personalized goals are perceived to be improving general well-being	"I think being able to set goals, like exercise for half an hour a day or sleep before 11 PM, might promote healthy living." (Participant #18, doctor)
Guidance (ie, provision of practical advice)		
Artificial intelligence–based chatbot	Loss of human contact for emotional support is deemed meaningless	"If you are having some emotional issues, you probably wouldn't want to talk to a robot who might not even understand your question properly." (Participant #21, nurse)
Chat with counselor	In-app counseling providing greater convenience and access to care compared to conventional mode of interaction	"I think the chat with a counselor might be helpful, as it is pretty hard to get an appointment with them. And with the built-in feature, one can get a response almost immediately." (Participant #17, doctor)
Chat with counselor	Enabling anonymous care	"I think we do a lot of texting nowadays, so I think texting with a counselor might help, especially if the counselor does not know our identity. So privacy is guaranteed while getting your problems sorted out with a professional." (Participant #34, nurse)
Framing (ie, shaping of information that alters its perceived nature)		
Personalized messages	Personalized messages offer a sense of human touch	"A message like this [personalized message] might encourage me to check on myself more frequently. It feels like someone is concerned and telling me to take care of myself." (Participant #12, doctor)
Social influence (ie, social support and conforming to a social trend)		
Forum	Emphasis on anonymity for concerns about risk of emotional manipulation and privacy	"Although it may be useful for sharing experiences and support one another, we will definitely need moderators to watch out for potential bullying or manipulation on the forum. In a support group, you can still have other people playing the [role] as a whistleblower, but for an anonymized forum chat, there may be someone who will keep posting malicious comments." (Participant #31, nurse)
App-based peer support group	Peer support perceived as useful as a platform for seeking comfort and sharing experience with colleagues	"For people who require emotional support but choose not to share their problems in a face-to-face format, I think this is a handy feature to have." (Participant #26, nurse)
Gamification (ie, tapping into an individual's desire to progress in a game setting)		
Interactive game activities	Reservations about the effectiveness of game elements in improving wellness	"I think it might motivate people who like to play games or have the habit of playing games regularly to use the app. But it might not appeal to people who do not play games." (Participant #40, doctor)
Scoring system	Comparing scores between users may discourage usage	"Personally, I feel that comparing points [between users] will only add to the stress, because people become competitive." (Participant #27, nurse)
App-based resources		
Mindfulness-based exercises	In-app tutorials and short readings perceived as handy and useful	"This [mindfulness-based exercise] is very useful. It can potentially help me destress myself after my shift and sleep well." (Participant #29, nurse)
Self-help aids (eg, short articles and videos)	In-app tutorials and short readings perceived as handy and useful	"Some people may want to look for useful articles, and if the articles are provided in the app itself, it will be really convenient and can potentially benefit a lot of people. It would be good if they are [presented] in point form." (Participant #40, doctor)

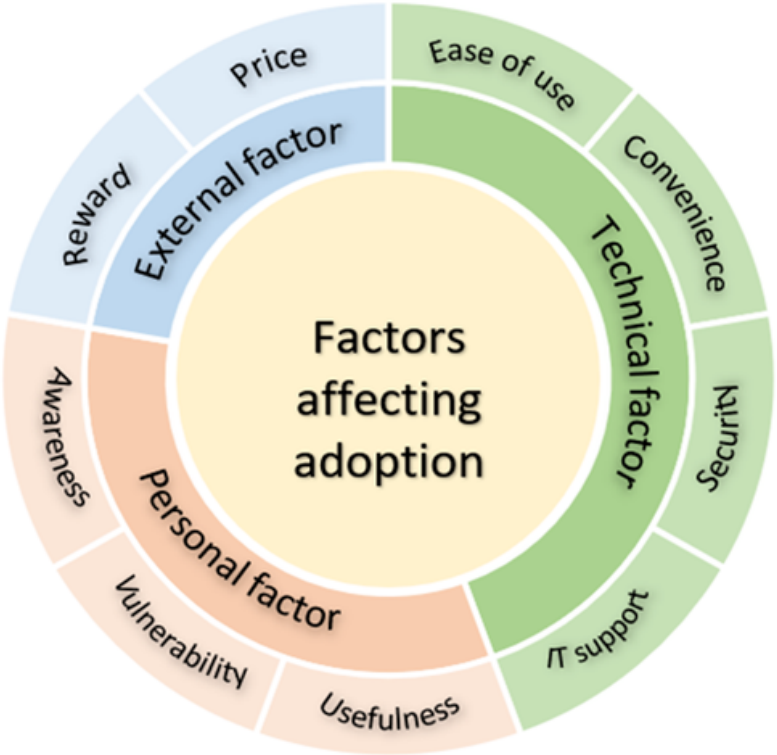
XSL•FO
RenderX

Factors That Could Affect Adoption and Sustained Use of a Wellness mHealth App

Factors that could potentially influence the adoption of a wellness mHealth app among frontline health care workers are presented in Table 4. Three overarching factors and nine themes were identified: (1) technical factors, with the themes perceived ease of use, convenience, security, and information technology (IT) support; (2) personal factors, with the themes perceived usefulness, perceived vulnerability, and awareness; and (3) external factors, with the themes rewards and price of app. Participants noted that nonintrusiveness with a minimum set of features relevant to the needs of frontline workers would be of prime importance in prompting adoption and ensuring continued use. As one participant noted, a simple interface that requires "minimum manual input" (Participant #12, doctor) was one of the most desired features for continued use. A robust IT and security system was also brought up by participants to help them start or continue using an mHealth app.

A few participants did not consider an mHealth app as necessary for promoting their mental health and wellness, mainly because they felt that they were coping well. Nonetheless, they thought that such an app might be helpful for colleagues who require psycho-emotional support or who are keen to improve their well-being. Increasing awareness of the availability of an app would be vital to its adoption. To encourage and sustain usage, affordable pricing of the app and attractive rewards could be considered to motivate users. Cost was one of the determining factors for app adoption: the majority of participants favored a free app or were willing to pay a small amount (eg, US $10) if it lived up to their expectations. More tangible and extrinsic rewards, such as Continuing Professional Education (CPE) points and taxi credits after accruing virtual points for adherence, were also suggested as motivation to sustain usage. Figure 1 summarizes the factors that may influence adoption and sustainability of a wellness mHealth app for frontline workers.

Figure 1. Factors affecting adoption and sustained use of a mobile health app. IT: information technology.



XSL•FO
RenderX

Table 4. Factors affecting potential adoption and sustained use of wellness mobile health (mHealth) apps.

Factors and themes	Illustrative quotes
Technical factors	
Perceived ease of use	"I will use the app if it is easy to use. I think judicious use of color might catch people's attention." (Participant #12, doctor)
Convenience	"Because we are always using the phone, if you are feeling down, at least you can use the app to help you manage your emotions. I think it is quite convenient and handy." (Participant #20, nurse)
Security and privacy	"A moderator has to be present to watch out for potential bullying and malicious comments [in forum chats] so as to make people feel safe." (Participant #34, nurse)
Information technology (IT) support	"It will be good if there is a team of dedicated IT personnel to ensure the app is running well and to minimize the potential [hacking] and leaking of information." (Participant #34, nurse)
Personal factors	
Tangible benefits	"I guess for the general health care workers, it [the app] might be useful in helping them to manage their emotions better, as there are many features that I feel are quite useful, such as the virtual counseling service and peer support." (Participant #41, doctor)
Perceived vulnerability to wellness risks	"I might not use the app routinely, as I have been coping well, but for some colleagues, I can tell they are struggling with burnout. If the app is introduced to them, I feel they are likely to use it." (Participant #19, nurse)
Awareness of app	"Many people did not participate in the wellness program, simply because they were not aware or did not feel the need to. Same thing for the app; it has to be made known to us before we can decide to use it." (Participant #26, nurse)
External factors	
Rewards and incentives	"I think you need some rewards to encourage usage, apart from CPE [Continuing Professional Education] points; I think it will be great if rewards like grocery vouchers and taxi credits can be exchanged with the points we have accumulated in the app." (Participant #21, nurse)
Cost	"I stopped using [previous mHealth app] because they wanted to charge a lot. So, if your app is free, then it will be wonderful. I will totally use it!" (Participant #5, doctor)

Discussion

Principal Findings

This study sought to build on limited literature by exploring the perceived usefulness of mHealth apps and features to improve psychosocial well-being from the perspectives of frontline workers responding to the COVID-19 pandemic. Previous studies have suggested that psychological interventions could help frontline workers cope with emotional exhaustion and could ameliorate work-induced stress during public health emergencies [8,38,39] and that counseling reduces the risk of depression in frontline workers [40]. However, our study found that our current in-person counseling has limitations in its ability to monitor and manage physical and mental health issues faced by frontline workers. Our participants, who are working on the front line, commonly reported that lack of rapport and limited continuity and awareness of the program inhibited uptake and appreciation of in-person counseling and psycho-emotional education. This finding from our study reinforces that from a previous study that insufficient rapport-building led to diminished utilization of wellness support services among frontline workers [31]. These factors, together, underscore the importance of developing an optimal set of psychosocial interventions that address pertinent concerns and needs of frontline workers. Delivery of psychological interventions through digital devices can possibly help in overcoming some of the barriers identified, with the potential for wide implementation in a convenient, efficient, and cost-effective manner [41]. Current evidence indicates that mHealth-based interventions increased accessibility to counseling and psycho-education and improved resilience among health care workers [42,43]. Leveraging digital platforms to provide wellness support to frontline workers may be a viable approach. Consistent with previous research, frontline workers generally had limited ability to recognize their psychological problems [44]. Constant exposure to a high-intensity work environment and chronic fatigue appeared to impair the participants' ability to recognize their psycho-emotional signs and symptoms. Hence, assisting frontline workers to identify their symptoms early in the process could be an important strategy for mitigating the risk of developing severe mental health issues. To this end, it is encouraging to note that frontline workers in our study desired a *reminder* functionality, such as tracking of mood and psychological stress, to understand their own mental health states. Another mHealth app feature that was highly valued by the frontline workers was in-app *resources*. This finding is consistent with evidence from previous studies indicating that reminders in mHealth apps played a vital role in improving emotional well-being. For instance, regular monitoring of mood was found to be associated with timely detection and mitigation of potential stressors [45]. Routine in-app reporting of thoughts and feelings can increase emotional self-awareness, which in turn, fosters coping skills [46]. A recent randomized controlled trial demonstrated that mood tracking resulted in improved coping ability, better adaptive response, and enhanced self-care [47]. Likewise, in-app resources, such as meditation and

XSL•FO
RenderX

mindfulness exercises, were found to have a positive effect on improving subjective well-being and behavioral regulation [48]. Therefore, future mHealth interventions should incorporate key nudge strategies that are favored by frontline workers into the design to increase effectiveness.

It has been well-recognized that social support, such as being esteemed, valued, and part of a social network of mutual assistance, is strongly associated with improved mental health [49-51]. Perceived availability of social support may bolster one's ability to cope with challenges, providing an avenue for emotional expression. Our study found that informal help-seeking from family and friends or non–help-seeking (ie, ignoring symptoms) were prevailing practices for frontline workers in response to emotional exhaustion associated with COVID-19. Frontline workers commonly reported the need for ongoing social support and desired an mHealth app with features that would enable them to readily access and check in with a virtual peer support community. Indeed, having similar experiences during the pandemic, peer health care workers may be in a unique position to provide practical and emotional support. A systematic review showed that digital peer support interventions led to improved self-coping and adaptation [27]. Through candid sharing of experiences and validation of feelings, app-based peer support, embedded in a *social influence* nudge strategy, could complement existing in-person support and have a positive impact on frontline workers' psychological well-being [52]. There was, however, a common concern among our participants regarding potential risks arising from virtual peer interactions, such as the spread of misleading information, derogatory comments, and disclosure of personal mental health issues. Therefore, an appropriate safeguarding mechanism should be in place to ensure the best possible outcomes for frontline workers.

Despite several mHealth features that were valued by frontline workers, a crucial issue remains to be considered: adoption and sustained engagement. A recent study reported that nearly 70% of users abandoned mHealth apps after a single use or stopped using the apps after a month [53]. This trend could stem from various factors. Participants in our study reported that mHealth app technicality and intrinsic factors were critical for uptake and retention. This finding resonates with prior research [54] and highlights that a simple user interface would be one of the most important aspects of an mHealth wellness program for frontline workers. Communicating the benefits and relevance of the wellness app, as well as transparency of data management, would be equally important for the successful implementation of an mHealth app [55,56]. In contrast to prior literature, frontline workers were more inclined to engage with a wellness app if external incentives, such as CPE points, were offered [57]. Therefore, it may be helpful to consider incorporating app features that can be manipulated to optimize incentive effectiveness (ie, type of incentives, timing, and magnitude) to maintain engagement of frontline workers.

Strengths and Limitations

This study added important evidence to the potential for wellness interventions delivered through an mHealth app from the perspectives of frontline health care workers. Findings from this study could provide valuable insight into the development and implementation of mHealth apps for improving the well-being of frontline workers involved in the response to COVID-19. Notwithstanding its strengths, this study was limited in several aspects. The results of this study were derived from qualitative research alone, which is, by nature, prone to a degree of potential subjectivity. Despite our efforts to recruit a balanced mix of genders, almost three-quarters of the participants were female. This might have introduced potential selection bias into the study. However, we did not find any considerable gender differences in viewpoints on key topics of interest. The discussion on mHealth features during the interviews was primarily based on visual materials underpinned by nudge strategies. Thus, participants' responses might have been influenced by the mHealth features presented. Lastly, the study participants were limited to doctors and nurses; further research is needed to explore perceptions of mHealth app features to support well-being among administrative and ancillary staff as well as allied health care professionals working on the front line.

Conclusions

Emerging technologies hold considerable promise for significantly expanding the reach of wellness programs for frontline health care workers. Traditional face-to-face modes of episodic support in times of public health emergencies seem to have limited utility to monitor and address the needs of frontline workers in a timely and holistic manner. Our study highlighted the need to take into account frontline workers' preferences and values when designing mHealth-based models of care to promote their well-being. Future work should focus on developing a nonintrusive and personalized app with in-app counseling and peer support features to improve well-being as well as tangible and extrinsic rewards to foster continued use.

Acknowledgments

This work was supported by a National University of Singapore COVID-19 Research Seed Grant (grant NUSCOVID19RG-29).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.

XSL•FO
RenderX

[DOC File , 68 KB - jmir_v23i5e26282_app1.doc]

Multimedia Appendix 2
Composition of participants in each session of one-on-one interviews and focus group discussions.
[DOCX File , 23 KB - jmir_v23i5e26282_app2.docx]

References

1. Armitage R, Nellums LB. Protecting health workers' mental health during COVID-19. Public Health 2020 Aug;185:18 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.044] [Medline: 32516622]
2. Spoorthy MS, Pratapa SK, Mahant S. Mental health problems faced by healthcare workers due to the COVID-19 pandemic-A review. Asian J Psychiatr 2020 Jun;51:102119 [FREE Full text] [doi: 10.1016/j.ajp.2020.102119] [Medline: 32339895]
3. Lehmann M, Bruenahl CA, Löwe B, Addo MM, Schmiedel S, Lohse AW, et al. Ebola and psychological stress of health care professionals. Emerg Infect Dis 2015 May;21(5):913-914 [FREE Full text] [doi: 10.3201/eid2105.141988] [Medline: 25897490]
4. Preti E, Di Mattei V, Perego G, Ferrari F, Mazzetti M, Taranto P, et al. The psychological impact of epidemic and pandemic outbreaks on healthcare workers: Rapid review of the evidence. Curr Psychiatry Rep 2020 Jul 10;22(8):43 [FREE Full text] [doi: 10.1007/s11920-020-01166-z] [Medline: 32651717]
5. Corley A, Hammond NE, Fraser JF. The experiences of health care workers employed in an Australian intensive care unit during the H1N1 influenza pandemic of 2009: A phenomenological study. Int J Nurs Stud 2010 May;47(5):577-585 [FREE Full text] [doi: 10.1016/j.ijnurstu.2009.11.015] [Medline: 20036360]
6. Greenberg N. Mental health of health-care workers in the COVID-19 era. Nat Rev Nephrol 2020 Aug;16(8):425-426 [FREE Full text] [doi: 10.1038/s41581-020-0314-5] [Medline: 32561871]
7. Sheraton M, Deo N, Dutt T, Surani S, Hall-Flavin D, Kashyap R. Psychological effects of the COVID 19 pandemic on healthcare workers globally: A systematic review. Psychiatry Res 2020 Oct;292:113360 [FREE Full text] [doi: 10.1016/j.psychres.2020.113360] [Medline: 32771837]
8. Tomlin J, Dalgleish-Warburton B, Lamph G. Psychosocial support for healthcare workers during the COVID-19 pandemic. Front Psychol 2020;11:1960 [FREE Full text] [doi: 10.3389/fpsyg.2020.01960] [Medline: 32849149]
9. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e15-e16. [doi: 10.1016/s2215-0366(20)30078-x]
10. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):300-302 [FREE Full text] [doi: 10.1016/S2215-0366(20)30073-0] [Medline: 32085840]
11. Banjar WM, Alaqeel MK. Healthcare worker's mental health dilemma during COVID-19 pandemic: A reflection on the KSA experience. J Taibah Univ Med Sci 2020 Aug;15(4):255-257 [FREE Full text] [doi: 10.1016/j.jtumed.2020.06.006] [Medline: 32837506]
12. Ripp J, Peccoralo L, Charney D. Attending to the emotional well-being of the health care workforce in a New York City health system during the COVID-19 pandemic. Acad Med 2020 Aug;95(8):1136-1139 [FREE Full text] [doi: 10.1097/ACM.0000000000003414] [Medline: 32282344]
13. Chidambaram S, Erridge S, Kinross J, Purkayastha S. Observational study of UK mobile health apps for COVID-19. Lancet Digit Health 2020 Aug;2(8):e388-e390. [doi: 10.1016/s2589-7500(20)30144-8]
14. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
15. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: A systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
16. Coronavirus pandemic prompts a surge in the number of people using NHS tech in 2020. NHS Digital. London, UK: National Health Service; 2020 Dec 21. URL: https://digital.nhs.uk/news-and-events/latest-news/surge-in-people-using-nhs-tech-2020 [accessed 2021-05-27]
17. Findlay S, Palma S, Milne R. Coronavirus contact-tracing apps struggle to make an impact. Financial Times. 2020 May 18. URL: https://www.ft.com/content/21e438a6-32f2-43b9-b843-61b819a427aa [accessed 2021-05-26]
18. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
19. Voyer B. 'Nudging' behaviours in healthcare: Insights from behavioural economics. Br J Healthc Manage 2015 Mar 02;21(3):130-135. [doi: 10.12968/bjhc.2015.21.3.130]
20. Xu Z, Geng J, Zhang S, Zhang K, Yang L, Li J, et al. A mobile-based intervention for dietary behavior and physical activity change in individuals at high risk for type 2 diabetes mellitus: Randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 03;8(11):e19869 [FREE Full text] [doi: 10.2196/19869] [Medline: 33141092]

XSL•FO
RenderX

21. Wiecek E, Torres-Robles A, Cutler RL, Benrimoj SI, Garcia-Cardenas V. Impact of a multicomponent digital therapeutic mobile app on medication adherence in patients with chronic conditions: Retrospective analysis. J Med Internet Res 2020 Aug 12;22(8):e17834 [FREE Full text] [doi: 10.2196/17834] [Medline: 32784183]
22. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: Iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
23. van Roy R, Zaman B. Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In: Ma M, Oikonomou A, editors. Serious Games and Edutainment Applications. Cham, Switzerland: Springer; 2017:485-509.
24. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: Qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
25. 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]
26. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
27. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CRC, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: Randomized controlled trial. J Med Internet Res 2019 Nov 07;21(11):e14269 [FREE Full text] [doi: 10.2196/14269] [Medline: 31697244]
28. Khalik S. Surge in coronavirus infections strains healthcare system. The Straits Times. 2020 Apr 18. URL: https://www.straitstimes.com/singapore/surge-in-virus-infections-strains-healthcare-system [accessed 2021-05-26]
29. Zhang LM. Circuit breaker to be lifted, Singapore to reopen gradually in 3 phases. The Straits Times. 2020 May 20. URL: https://www.straitstimes.com/singapore/circuit-breaker-to-be-lifted-singapore-to-reopen-gradually-in-3-phases [accessed 2021-05-26]
30. Tomlin J, Dalgleish-Warburton B, Lamph G. Psychosocial support for healthcare workers during the COVID-19 pandemic. Front Psychol 2020;11:1960 [FREE Full text] [doi: 10.3389/fpsyg.2020.01960] [Medline: 32849149]
31. Muller A, Hafstad E, Himmels J, 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]
32. Si M, Su X, Jiang Y, Wang W, Gu X, Ma L, et al. Psychological impact of COVID-19 on medical care workers in China. Infect Dis Poverty 2020 Aug 12;9(1):113 [FREE Full text] [doi: 10.1186/s40249-020-00724-0] [Medline: 32787929]
33. Sittig S, Hauff C, Graves RJ, Williams SG, McDermott RC, Fruh S, et al. Characteristics of and factors influencing college nursing students' willingness to utilize mHealth for health promotion. Comput Inform Nurs 2020 May;38(5):246-255 [FREE Full text] [doi: 10.1097/CIN.0000000000000600] [Medline: 32032084]
34. Kwan YH, Cheng TY, Yoon S, Ho LYC, Huang CW, Chew EH, et al. A systematic review of nudge theories and strategies used to influence adult health behaviour and outcome in diabetes management. Diabetes Metab 2020 Nov;46(6):450-460 [FREE Full text] [doi: 10.1016/j.diabet.2020.04.002] [Medline: 32387700]
35. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Abingdon, UK: Routledge; 2017.
36. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd edition. Thousand Oaks, CA: SAGE Publications, Inc; 1998.
37. 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]
38. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e15-e16 [FREE Full text] [doi: 10.1016/S2215-0366(20)30078-X] [Medline: 32085839]
39. Vesel L, Waller K, Dowden J, Fotso JC. Psychosocial support and resilience building among health workers in Sierra Leone: Interrelations between coping skills, stress levels, and interpersonal relationships. BMC Health Serv Res 2015;15 Suppl 1:S3 [FREE Full text] [doi: 10.1186/1472-6963-15-S1-S3] [Medline: 26062735]
40. Bower P, Knowles S, Coventry PA, Rowland N. Counselling for mental health and psychosocial problems in primary care. Cochrane Database Syst Rev 2011 Sep 07(9):CD001025 [FREE Full text] [doi: 10.1002/14651858.CD001025.pub3] [Medline: 21901675]
41. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
42. White A, Thomas DSK, Ezeanochie N, Bull S. Health worker mHealth utilization: A systematic review. Comput Inform Nurs 2016 Mar 7:5-12. [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]

XSL•FO
RenderX

43. Ventola CL. Mobile devices and apps for health care professionals: Uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
44. Styra R, Hawryluck L, Robinson S, Kasapinovic S, Fones C, Gold WL. Impact on health care workers employed in high-risk areas during the Toronto SARS outbreak. J Psychosom Res 2008 Feb;64(2):177-183 [FREE Full text] [doi: 10.1016/j.jpsychores.2007.07.015] [Medline: 18222131]
45. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: Understanding expert perspectives and student experiences. In: Proceedings of the 35th SIGCHI Conference on Human Factors in Computing Systems. 2017 Presented at: 35th SIGCHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 629-641 URL: http://europepmc.org/abstract/MED/28920106 [doi: 10.1145/3025453.3025750]
46. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: Review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
47. 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]
48. Janssen M, Heerkens Y, Kuijer W, van der Heijden B, Engels J. Effects of mindfulness-based stress reduction on employees' mental health: A systematic review. PLoS One 2018;13(1):e0191332 [FREE Full text] [doi: 10.1371/journal.pone.0191332] [Medline: 29364935]
49. Harandi TF, Taghinasab MM, Nayeri TD. The correlation of social support with mental health: A meta-analysis. Electron Physician 2017 Sep;9(9):5212-5222 [FREE Full text] [doi: 10.19082/5212] [Medline: 29038699]
50. Ozbay F, Johnson DC, Dimoulas E, Morgan CA, Charney D, Southwick S. Social support and resilience to stress: From neurobiology to clinical practice. Psychiatry (Edgmont) 2007 May;4(5):35-40 [FREE Full text] [Medline: 20806028]
51. Taleghani F, Babazadeh S, Mosavi S, Tavazohi H. The effects of peer support group on promoting quality of life in patients with breast cancer. Iran J Nurs Midwifery Res 2012 Feb;17(2 Suppl 1):S125-S130 [FREE Full text] [Medline: 23833594]
52. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: Systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
53. 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]
54. 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]
55. Holden RJ, Karsh B. The technology acceptance model: Its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
56. To W, Lee PKC, Lu J, Wang J, Yang Y, Yu Q. What motivates Chinese young adults to use mHealth? Healthcare (Basel) 2019 Dec 02;7(4):156 [FREE Full text] [doi: 10.3390/healthcare7040156] [Medline: 31810293]
57. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mHealth app: Process evaluation of the Carrot Rewards app. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
CPE: Continuing Professional Education
IT: information technology
mHealth: mobile health
NHS: National Health Service

Edited by R Kukafka; submitted 05.12.20; peer-reviewed by YH Kwan, S Sittig; comments to author 24.12.20; revised version received 13.01.21; accepted 19.04.21; published 31.05.21.

Please cite as:
Yoon S, Goh H, Nadarajan GD, Sung S, Teo I, Lee J, Ong MEH, Graves N, Teo TL
Perceptions of Mobile Health Apps and Features to Support Psychosocial Well-being Among Frontline Health Care Workers Involved in the COVID-19 Pandemic Response: Qualitative Study
J Med Internet Res 2021;23(5):e26282
URL: https://www.jmir.org/2021/5/e26282
doi:10.2196/26282
PMID:33979296

XSL•FO
RenderX

©Sungwon Yoon, Hendra Goh, Gayathri Devi Nadarajan, Sharon Sung, Irene Teo, Jungup Lee, Marcus E H Ong, Nicholas Graves, Tess Lin Teo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2021. 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

Pain Recognition With Electrocardiographic Features in Postoperative Patients: Method Validation Study

Emad Kasaeyan Naeini[1*], BSc, MSc; Ajan Subramanian[1*], BSc; Michael-David Calderon[2], BSc; Kai Zheng[3], PhD; Nikil Dutt[1], PhD; Pasi Liljeberg[4], PhD; Sanna Salantera[5,6], PhD, RN; Ariana M Nelson[7], MD; Amir M Rahmani[1,8,9], PhD

[1]Department of Computer Science, University of California, Irvine, Irvine, CA, United States
[2]Department of Anesthesiology and Perioperative Care, UC Irvine Health, Orange, CA, United States
[3]Department of Informatics, University of California, Irvine, Irvine, CA, United States
[4]Department of Future Technology, University of Turku, Turku, Finland
[5]Department of Nursing, University of Turku, Turku, Finland
[6]Turku University Hospital, Turku, Finland
[7]School of Medicine, University of California, Irvine, Irvine, CA, United States
[8]School of Nursing, University of California, Irvine, Irvine, CA, United States
[9]Institute for Future Health, University of California, Irvine, Irvine, CA, United States
[*]these authors contributed equally

Corresponding Author:
Emad Kasaeyan Naeini, BSc, MSc
Department of Computer Science
University of California, Irvine
DBH Bldg, 3rd Fl.
Irvine, CA, 92617
United States
Phone: 1 9496106661
Email: ekasaeya@uci.edu

Abstract

Background: There is a strong demand for an accurate and objective means of assessing acute pain among hospitalized patients to help clinicians provide pain medications at a proper dosage and in a timely manner. Heart rate variability (HRV) comprises changes in the time intervals between consecutive heartbeats, which can be measured through acquisition and interpretation of electrocardiography (ECG) captured from bedside monitors or wearable devices. As increased sympathetic activity affects the HRV, an index of autonomic regulation of heart rate, ultra–short-term HRV analysis can provide a reliable source of information for acute pain monitoring. In this study, widely used HRV time and frequency domain measurements are used in acute pain assessments among postoperative patients. The existing approaches have only focused on stimulated pain in healthy subjects, whereas, to the best of our knowledge, there is no work in the literature building models using real pain data and on postoperative patients.

Objective: The objective of our study was to develop and evaluate an automatic and adaptable pain assessment algorithm based on ECG features for assessing acute pain in postoperative patients likely experiencing mild to moderate pain.

Methods: The study used a prospective observational design. The sample consisted of 25 patient participants aged 18 to 65 years. In part 1 of the study, a transcutaneous electrical nerve stimulation unit was employed to obtain baseline discomfort thresholds for the patients. In part 2, a multichannel biosignal acquisition device was used as patients were engaging in non-noxious activities. At all times, pain intensity was measured using patient self-reports based on the Numerical Rating Scale. A weak supervision framework was inherited for rapid training data creation. The collected labels were then transformed from 11 intensity levels to 5 intensity levels. Prediction models were developed using 5 different machine learning methods. Mean prediction accuracy was calculated using leave-one-out cross-validation. We compared the performance of these models with the results from a previously published research study.

Results: Five different machine learning algorithms were applied to perform a binary classification of baseline (BL) versus 4 distinct pain levels (PL1 through PL4). The highest validation accuracy using 3 time domain HRV features from a BioVid research

XSL•FO
RenderX

paper for baseline versus any other pain level was achieved by support vector machine (SVM) with 62.72% (BL vs PL4) to 84.14% (BL vs PL2). Similar results were achieved for the top 8 features based on the Gini index using the SVM method, with an accuracy ranging from 63.86% (BL vs PL4) to 84.79% (BL vs PL2).

Conclusions: We propose a novel pain assessment method for postoperative patients using ECG signal. Weak supervision applied for labeling and feature extraction improves the robustness of the approach. Our results show the viability of using a machine learning algorithm to accurately and objectively assess acute pain among hospitalized patients.

International Registered Report Identifier (IRRID): RR2-10.2196/17783

(J Med Internet Res 2021;23(5):e25079) doi:10.2196/25079

KEYWORDS

pain assessment; recognition; health monitoring; wearable electronics; machine learning

Introduction

Overview

Pain assessment is a critical public health burden and is essential to effective pain management, which is associated with many illnesses [1]. Pain is “an unpleasant sensory and emotional experience expressed in terms of actual or potential tissue damage” [2], according to the most widely accepted definition. Pain is considered to be a subjective experience that is related to each individual in early life through experiences related to injury [3]. Such pain, which is termed acute pain, usually lasts hours, days, or weeks. Acute pain is associated with soft tissue damage, a surgical procedure, or a brief disease process and fosters avoidance of the harmful action in the future and promotes healing by inhibiting activities that might cause further tissue damage [4]. Pain, as a susceptible and ambiguous phenomenon, is difficult to quantify [5], particularly when the patient's own opinion is difficult to reach due to their limited ability to communicate, as in patients under sedation or anesthesia, persons with intellectual disabilities, infants, and patients during critical illness [6]. Uncontrolled pain could cause some serious complications and may evolve into chronic pain. This could cause longer recovery in hospitals and delayed discharge, higher health care costs, and major psychological, financial, and social ramifications for patients [7]. However, overtreatment of pain can also result in adverse effects such as hospital readmission due to poorly controlled pain after discharge or long-term opioid dependence.

The current and “gold standard” pain assessment relies on patient self-reporting with tools such as the Visual Analogue Scale (VAS) and Numerical Rating Scale (NRS). Although these unidimensional models are considered powerful in acute pain assessment, they are rife with deficits given their interactive communication requirement between the patient and nurses, which is a serious problem in noncommunicative patients [8,9]. Such tools rely on nurses’ knowledge, physical assessment skills, and interviewing techniques. It is thus meaningful to develop better tools to assess pain intensity for continuous real-time pain monitoring. Such a tool not only improves the care process of noncommunicative patients but can also benefit other patient populations with timelier treatment, accurate assessment, and reduced monitoring burden on clinicians [10].

State-of-the-art objective pain intensity assessment algorithms consist of analysis of physiological and physical pain indicators, as multiparameter analysis is superior to a singular physiological parameter [11-13]. Objective pain assessment leverages using wearable devices to capture the physiological parameters. Internet-of-Things (IoT) devices, including wearable devices, play a significant role in objective pain monitoring systems [10]. As an example, Vatankhah et al [14] measured and diagnosed pain levels of human using discrete wavelet transform via electroencephalographic signals. These devices are in charge of various real-time health monitoring services as well as continuously processing and analyzing pain intensity levels to classify them automatically and objectively. However, these solutions to date have only been evaluated on healthy volunteers (stimulated pain). This was the motivation to develop a multimodal data set from postoperative adult patients in hospitals to get a better understanding of pain intensity characteristics of real patients. We call the data set UCI iHurt Database (UCI_iHurtDB) [15]. The data set is planned to be released for research purposes and, to the best of our knowledge, is the first comprehensive data set collected from patients suffering from real postoperative pain.

Electrocardiography (ECG) is useful for indicating the perception of pain among all of the physiological signals captured by wearable devices. Heart rate (HR) and heart rate variability (HRV) are the essential parameters that can be derived from ECG, as they are both coupled to autonomic nervous system activity, when internal body functions are involuntarily regulated, and they can provide a suitable proxy for examining pain intensity [16]. The most frequently used vital sign in pain studies is HR as the number of heartbeats, while HRV features with the extent to which the heart rate changes over a time interval or the extent to which it is spread over different frequencies are observed individually and selectively in some pain studies [17-19]. There are several approaches for classifying pain intensity of healthy subjects using machine learning techniques. Lopez-Martinez and Picard [18] explored traditional machine learning algorithms such as logistic regression and support vector machine (SVM) with different kernels, as well as recurrent neural network, to create a model for no pain versus pain at different intensities. Koenig et al [19] revealed that HRV is a promising measure of autonomic reactivity to nociceptive stimulation in healthy adults. Therefore, we examine HR and HRV features with their correlation with pain for future biosignal fusion in pain intensity assessment and pain detection.

XSL•FO
RenderX

To the best of our knowledge, this is the first work to study the relations between ECG physiological signal and pain intensity of postoperative patients to predict different pain intensity levels. This will promote advancements in both observational and physiological pain measurement. The technology used in this paper for objective pain assessment was developed by our group and presented in our previous work by Sarker et al [20]. The prototype device can be variably configured for inclusion in a wide range of applications. It is also validated on healthy adults under thermal and electrical experimental pain stimulus [21].

In this paper, we looked into the ECG signal as one of the physiological signals from the human body and examined HR and HRV features extracted from ECG in several categories to assess their correlation with pain intensity for future physiological data fusion in pain intensity assessment. Due to limited available self-report pain intensity labels, we explored weak supervision, which has shown that performance of the end-model can asymptotically improve with data set size, although noisy sources of supervision are used [22,23]. Thus, the contributions of this work are twofold: (1) We present UCI_iHurtDB, a freshly collected data set from postoperative patients consisting of multimodal biosignals (ECG, electromyography, electrodermal activity, photoplethysmography, accelerometer). (2) We provide a novel weakly supervised method to enhance the sparsely labeled data set. To the best of our knowledge, this is the first study using weak supervision in pain assessment.

Interpretation of ECG in Pain Studies

Interpretation of ECG in pain studies starts with the ability to detect the QRS complex as one of the morphological parts of the ECG waves, with the focus on the RR intervals (distance between adjacent R-waves). These peaks are essential in HRV analysis, where in literature they are usually referred to as NN intervals. HRV features have been traditionally calculated over a short period of 5 minutes or over a long period in 24 hours. However, in some cases such as acute physiological changes, some HRV features in less than a minute are also taken on as ultra–short-term analysis. Within each time window, HRV features can be extracted from NN intervals in several domains including (1) time domain features and (2) frequency domain features. HRV features in the time domain consist of statistical features including but not limited to SD of NN intervals, SD of average NN intervals, root mean square of successive differences (RMSSD), NN50, pNN50, and Max(HR) − min(HR). HRV features in the frequency domain are extracted by the Welch averaging method using power spectral density.

In literature, HRV analysis has been examined as one of the main ways to measure pain in different types; in Sesay et al [24], for instance, with the majority of 120 patients, it was observed that regarding acute pain after minor surgery, NRS was correlated with low-frequency (LF) band and the ratio of LF to high-frequency (HF) band but not with HF. To monitor the nociception level of patients with multiple physiological parameters, HF in a 1-minute window was calculated in Ben-Israel et al [11]. Jiang et al [17] experimented with the correlation of HRV features in the ultrashort term with acute pain. They suggested that multiple HRV features can indicate the change from no pain to pain. Werner et al [25] compared no pain among pain levels 1-4 using a random forest (RF) classifier. They reported the detection of pain using a set of features from ECG signals.

Methods

Setting

The study was approved by the University of California, Irvine (UCI) institutional review board (HS: 2017-3747). Candidates were selected from the Acute Pain Service (APS) patient list at UCI Health in Orange, California. The APS unit at the medical center serves approximately 100 patients weekly, enabling the lead Doctor of Medicine to recruit patients.

Study Description, Participants, and Recruitment

This study is a prospective observational data collection from postoperative patients likely having mild to moderate pain. All 25 participants recruited for this study met the following criteria: (1) age at least 18 years, (2) received a consult by the APS, (3) able to communicate, (4) able to provide written informed consent, and (5) healthy, intact facial skin. They were excluded if they had (1) any diagnosed condition affecting cognitive functions (dementia, psychosis), (2) any diagnosed condition affecting the central nervous system, facial nerves or muscles, (3) deformities on hand that prevent sensor placement, or (4) significant facial hair growth in the area where the sensors were going to be attached.

Potential participants were selected if they were determined to be eligible to participate in this study based on the aforementioned inclusion and exclusion criteria. Patients got both oral and written information about the details of the study. Candidates were provided at least 24 hours to consider participation in the study before finalizing the consent form, and they were recruited to participate in this study after obtaining the written consent form.

Study Design

After obtaining the written consent form, approximately 30 minutes of continuous ECG data was collected from the patients in their private room using multiple wearable sensors in two parts. In the first part, we used a transcutaneous electrical nerve stimulation (TENS) unit to obtain the baseline of the person. Patients were asked to increase the intensity of the TENS device up to the level that was tolerable for them, hold it for at least 10 seconds, and then decrease it to level 0, including additional rest between TENS challenges. In the second part, patients were engaged with soft activities such as walking, coughing, sitting, and lifting legs that may cause pain sensation. To achieve a better statistical analysis, data collection in both parts, with and without TENS unit, was repeated two to three times. Subject's self-report of pain was recorded using NRS. The NRS for pain is a segmented numeric version of the VAS to measure pain intensity in one dimension in which a respondent points to the number on the NRS, an integer from 0 to 10, that best represents their pain intensity from “no pain” to “worst pain” [26]. We expect to find solutions from multiple parameters that are robust in response to different acute pain cases or study designs.

XSL•FO
RenderX

Data Collection

To develop an algorithm for pain assessment in hospitalized patients, we tracked physiological signals such as HR, HRV, and respiratory rate from their ECG signals. The technology used in this study to capture the aforementioned signals includes the following components:

1. Biopotential acquisition system for ECG recording—ECG is a biopotential signal captured from the skin surface with a device developed by Sarker et al [20]. The system includes commercially available electrodes (eg, in 24 mm diameter), electrode-to-device lead wires, an ADS1299-based portable device, and computer software receiving streaming data from the portable device. The raw data of each channel at the rate of 500 samples per second is sent to the computer software through Bluetooth. The software visualizes the waveforms and saves the raw data into files. The common reference electrode is placed on the neutral bony area behind the ear. This device uses two channels to collect 2-lead ECG. One channel is to measure the potential between ECG - right arm and reference, and the other channel is to measure the potential between ECG - left arm and reference.
2. TENS unit device (Food and Drug Administration–cleared Class II over-the-counter HealthmateForever YK15AB electrotherapy device)—TENS units work by delivering small electrical impulses through electrodes that have adhesive pads to attach them to a person's skin.

Data Preprocessing

Data Synchronization

The ECG signals from each patient were sampled at a rate of 500 Hz. Data from two channels (left arm, right arm) were obtained. The patient's pain levels were simultaneously reported and saved as labels. For the purpose of synchronicity, the corresponding Unix timestamps were also obtained while extracting both ECG and label data. The ECG signals were trimmed from start to end to match the corresponding label timestamps. Since the sampling frequency was 500 Hz, each timestamp had 500 ECG samples associated with it.

Peak Detection

Once these clean signals were obtained, the second step in the pipeline was to extract peaks. To find the peaks, the signals were first sampled down to 250 Hz. A bidirectional long short-term memory network was used to obtain the probabilities and locations of peaks [27]. A window size of 1000 samples and stride of 100 samples was used to generate these predictions. Mean values were obtained from predictions that came from overlapping windows. The predictions that were below a particular threshold (0.05) were discarded and filtered out. Only those peaks that were in local maximum were selected. Once the peaks were obtained, the signal was resampled back to 500 Hz, and the peak probabilities and locations were obtained. This method, however, might still be susceptible to false positives that are likely generated due to the presence of noise or irregular heartbeats. Therefore, another preprocessing step that removes peaks that occur too close to each other was employed. A rolling window was used to remove peaks that occurred in a time period of 450 milliseconds or less between neighboring peaks. The final selected peaks were then appended with their corresponding Unix timestamps. This process was repeated for every patient.

Noise Removal

The third and final preprocessing step is to remove noise from the NN interval data. NN intervals are the time intervals between two successive peaks. They are obtained by subtracting two successive peak indices. All data points that are within 2 standard deviations of the mean were selected. The rest of the data points were considered outliers and were removed. Even after removing these outliers, however, there might still be anomalous (not a number [NaN]) values in noisy sections of the data. If the proportion of NaN values exceeded 50 percent, the noisy sections were discarded. Otherwise, only NaN values were discarded, and the remaining values were interpolated. The filtered NN intervals were then saved and used for feature extraction.

Feature Extraction

In the experiments conducted by Werner et al [25], they used 5.5 seconds of ECG signals and extracted 3 time domain HRV features from the ECG signal of each subject: (1) the arithmetic mean of time in between consecutive heartbeats or mean of NN intervals (AVNN), (2) RMSSD, and (3) the slope of the linear regression of NN intervals or the measure of acceleration of the heart rate. We use an open-source Python toolbox named pyHRV [28] to compute HRV features in our feature extraction process.

In addition to the 3 time domain features used by Werner et al [25], we also computed a few other time domain and frequency domain features. For classification, we conducted two separate experiments: (1) with the three above features only and (2) with additional time and frequency domain features using feature selection. The additional extracted features are mentioned below.

Time Domain Features

There were 19 time domain features extracted from the NN interval series. They include the slope of the ECG signals and 18 statistical measures. These features were computed using 5.5-second sliding windows. The definitions of these time domain features are mentioned in Table 1.

XSL•FO
RenderX

Table 1. Time domain heart rate variability features and their definitions.

Feature	Description
HR[a] (ms)	Beats per minute
AVNN (ms)	Mean of NN intervals
SDNN (ms)	Standard deviation of NN intervals
RMSSD (ms)	Root mean square of successive NN interval differences
NNXX (ms)	Number of NN interval differences greater than the specified threshold
pNNXX (%)	Percentage of successive NN intervals that differ by more than XX ms

[a]HR: heart rate.

The breakdown of the 19 aforementioned features is explained as follows: slope of NN intervals—a polynomial fit of degree 1; 5 NN interval features—total count, mean, minimum, maximum, SD; 9 NN interval difference features—mean difference, minimum difference, maximum difference, SD of successive interval differences, root mean square of successive interval differences, number of interval differences greater than (a) 20 milliseconds and (b) 50 milliseconds, and percentage of successive interval differences that differ by more than (a) 20 milliseconds and (b) 50 milliseconds; and 4 heart rate features—mean, minimum, maximum, and SD.

Frequency Domain Features

There were 13 frequency domain features extracted by estimating of power spectral density using the Welch method. These features were computed using 250-second rolling windows with a minimum threshold of 50 values per window. The definitions of these frequency domain features are mentioned in Table 2.

Table 2. Frequency domain heart rate variability features and their definitions.

Feature	Description
VLF power (s^2)	Absolute power in very low-frequency band (≤0.04)
LF power (s^2)	Absolute power in low-frequency band (0.04-0.15)
HF power (s^2)	Absolute power in high-frequency band (0.15-0.4)
LF peak (Hz)	Peak frequency in low-frequency band (0.04-0.15)
HF peak (Hz)	Peak frequency in high-frequency band (0.15-0.4)
Total power (s^2)	Total power over all frequency bands
LF/HF (%)	Ratio of LF-to-HF power

The breakdown of the 13 frequency domain features is explained as follows: total power—total spectral power over all frequency bands; 4 HF band fast Fourier transform (FFT) features—peak, absolute, relative, normalized; 4 LF band FFT features—peak, absolute, relative, normalized; 3 very low-frequency (VLF) band FFT features—peak, absolute, relative; and FFT ratio of HF and LF bands.

Feature Selection

To ensure generalization and avoid overfitting, it is important to perform feature selection. This, in turn, reduces computational complexity and the time for training and validating models. Feature selection models can be placed under three broad categories: filter-based methods, wrapper-based methods, and embedded methods. Filter-based methods statistically determine the relationship between input variables (features) and the target variable (label). They provide a metric for evaluating and filtering out features that will be used by the model. They are also computationally cheaper than the other two methods and have a reduced risk of overfitting [29]. Among the filter-based methods, Gini impurity/information gain is a widely used method to select the most informative features for a classification problem. Usually, a decision tree–based model like an RF classifier is used to output a feature importance vector. Every node of a decision tree represents a condition on how to split values present in a single feature. In this process, similar data on the condition variable end up on the same side of the split. The splitting condition is based on impurity of the features chosen in every node. During the training process, how much each feature contributed to the decrease in impurity is calculated, and features are then ranked based on this measure.

Labels

Label Distribution

Table 3 shows the distribution of NRS pain labels reported by patients during the clinical trials.

XSL•FO
RenderX

Table 3. Numerical Rating Scale distribution for 11 pain classes.

Reported Numerical Rating Scale labels	n
0	37
1	52
2	37
3	61
4	83
5	44
6	32
7	16
8	46
9	26
10	4

Since the NRS labels recorded during clinical trials were collected from real postoperative patients, there are some inherent challenges due to the distribution of data. For example, there are 83 occurrences of NRS pain label 4, but there are only 4 occurrences of NRS pain label 10 among all patients. Due to the subjective nature and the different sources of pain among our recruited patients, the imbalanced distribution of pain levels among all patients is inevitable.

To compare our pain assessment algorithm's performance with Werner et al [25], we downsampled our pain labels from 11 NRS classes (0-10) to 5 classes (0-4). Data points from NRS pain label 0 were considered as a baseline, and the remaining NRS pain labels were distributed among 4 classes. Thresholds for each downsampled class were carefully chosen in order to minimize an imbalanced class distribution. Table 4 shows the resulting distribution after downsampling the NRS pain labels. The relatively large number of occurrences of NRS pain label 4 increased the number of downsampled PL2 labels over other downsampled pain labels.

Table 4. Downsampled pain distribution with 5 classes.

Downsampled pain labels	n
BL[a]	37
PL1[b]	89
PL2	144
PL3	92
PL4	76

[a]BL: baseline.

[b]PL: pain level.

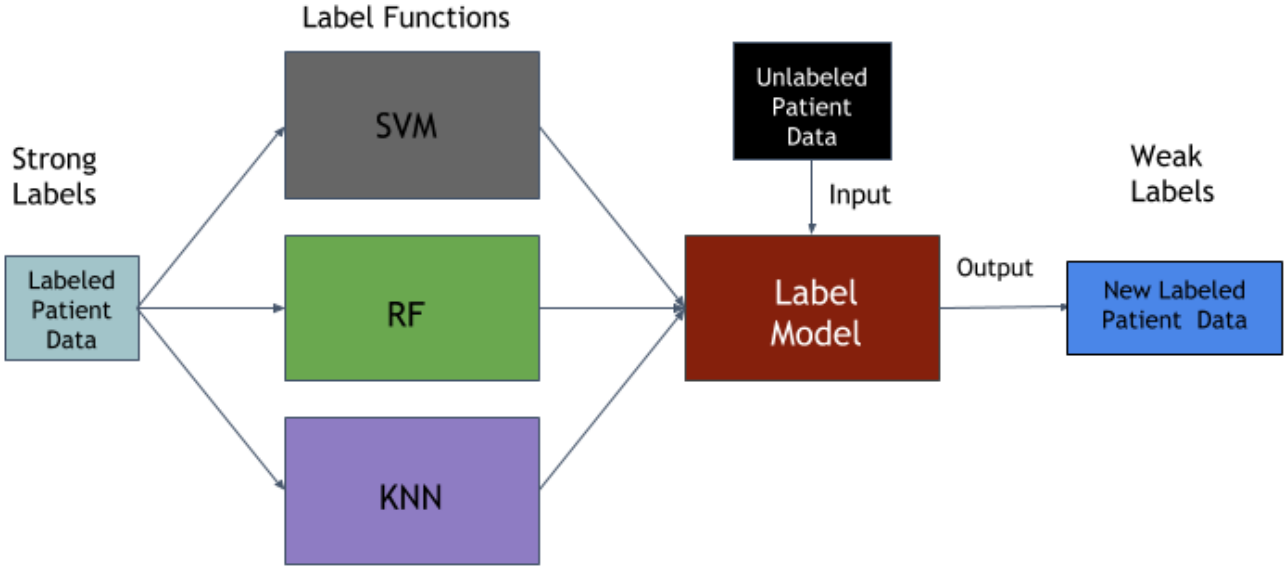
Labeling ECG Features

Since the patients' NRS values were only reported after performing some activities, labels were stored sparsely. While combining ECG features with their corresponding labels, their timestamps were matched based on the nearest 5.5 seconds (labeling threshold). More precisely, any ECG feature window that was within 5.5 seconds of a reported NRS value was given that value as its label. However, as a consequence of this, all the feature windows that were not within the labeling threshold were not given a corresponding label.

To label the remaining unlabeled data points, we employed Snorkel [23], a weak supervision framework for rapid training data creation. Snorkel is an end-to-end system that combines weak supervision sources to label training data with limited ground truth data. Rather than hand-labeling training data, Snorkel allows its users to write labeling functions that make use of heuristics, patterns, third-party machine learning models, external knowledge bases, and more. Weak supervision is typically employed to label large amounts of unlabeled data when there are noisy, limited, or imprecise sources. This approach eliminates the burden of continuously obtaining NRS values from patients. The use of Snorkel in our labeling process allowed us to make use of more data for the purpose of training and testing our pain algorithm. This consequently led to better performance during validation. Figure 1 depicts the architecture that was used to label these unlabeled instances.

XSL•FO
RenderX

Figure 1. Snorkel labeling architecture. KNN: k-nearest neighbor; RF: random forest; SVM: support vector machine.



All the data points that were within the labeling threshold were considered "strong" labels, or labels used for training labeling functions. These strong labels were collected directly from the patients. The remaining unlabeled data points were considered "weak" data and were kept aside for the weak supervision algorithm to label. For the labeling process, only each patient's strongly labeled data was used to label their own unlabeled instances. This was done to avoid the possibility of data leakage during the validation process.

The labeling functions consist of a group of three off-the-shelf machine learning models: (1) an SVM with a radial basis function kernel, (2) an RF classifier, and (3) a k-nearest neighbor (KNN) classifier with uniform weights. Once each of these models is trained on strong labels, they are used to make predictions on the weak or unlabeled data. Their predictions are then collected and converted into a single confidence-weighted label per data point using Snorkel's *LabelModel* function. The most confident label predictions from each datapoint were considered as labels for the weak data. It is important to note that weak supervision does not compromise on the reliability of our algorithm because we use the weakly labeled data only for training our models. This way, the performance of our algorithm can be measured using only real data collected from patients.

Classification

To compare the performance of our pain algorithm with Werner et al [25], we performed binary classification on our test data using the 3 time domain features mentioned in their work. We split the binary classification problem into 4 different categories: baseline (BL) versus pain level 1 (PL1), BL versus PL2, BL versus PL3, and BL versus PL4. Since 1 of the patients had data from only one downsampled label class, they were discarded from the classification process. Consequently, we were left with data from 19 patients.

We evaluated the performance of our pain algorithm using leave-one-out cross-validation (LOOCV) with the focus on optimizing the area under the curve (AUC) score. During each iteration of LOOCV, the data of 18 of the 19 patients, including those data points that were labeled by Snorkel, were used for training. For testing, only the strongly labeled data points from the one patient left out were used. This process is repeated for all 19 patients to estimate the algorithm's performance on unseen data. Due to the presence of an imbalanced distribution of pain levels within patients, data points from some pain levels were nonexistent from their data. As a result, it was not possible to compute either precision or recall for most patients.

The following five classification methods were deployed in our experiments to identify the best performing model for our pain assessment algorithm: AdaBoost classifier, XGBoost classifier, RF classifier, SVM classifier, and KNN classifier.

We also conducted separate experiments with feature selection using the 32 features mentioned in the Feature Extraction section. To get the best set of features for classification, we run LOOCV using an RF classifier. We compute the Gini importance of each of the features at every fold and select those features that were at least one standard deviation above the mean importance score. As a result, it was possible to have different sets of features in every fold. After computing the best set of features at every fold, we consider those features that were used in most of the folds for classification. The following 8 features were used in the final feature set: (1) total spectral power, (2) absolute LF power, (3) absolute HF power, (4) mean HR, (5) relative HF power, (6) normalized HF power, (7) relative VLF power, and (8) normalized LF power.

Results

Pain Demographic Characteristics

A total of 25 patients with acute pain were engaged by APS and recruited for this study at UCI Medical Center. However, the ECG data from 2 patients were missing due to connectivity issues. Moreover, we found that 3 of the patients had arrhythmia, so we removed those 3 as well. The average age of patients was 55.6 years (SD 16.24, range 23-89); 52% (13/25) of patients were male and 48% (12/25) of patients were female (Table 5). All of the patients (n=20) were taking prescription medication

XSL•FO
RenderX

at the time of the study. The nature of the procedures for each participant included the following domains: 50% general surgery (diagnostic laparoscopy, exploratory laparotomy, and vascular), 25% orthopedics, 15% trauma (thoracic pain and rib plating), and 10% urology (cystectomy and bladder augmentation).

Table 5. Patient demographic characteristics (N=25).

Variable	Value	Range
Patients excluded due to arrhythmia, n (%)	3 (12)	N/A[a]
Patients excluded due to missing ECG[b] data, n (%)	2 (8)	N/A
Gender, male, n (%)	13 (52)	N/A
Weight (kg), mean (SD)	76.56 (17.31)	52.2-112.2
Height (cm), mean (SD)	170.9 (10.44)	152.4-193
BMI[c] (kg/m^2), mean (SD)	26.33 (6.14)	15.1-38.73
Procedure domain (n=20), n (%)		
General surgery	10 (50)	N/A
Orthopedics	5 (25)	N/A
Trauma	3 (15)	N/A
Urology	2 (10)	N/A

[a]N/A: not applicable.

[b]ECG: electrocardiography.

[c]BMI: body mass index.

Pain Engagement Results

To make a fair comparison between our pain assessment algorithm and the work of Werner et al [25], we replicated their settings into our data set. The comparisons of the accuracy achieved by our algorithm on all five classifiers while using only 3 time domain features are shown in Figure 2. Similarly, Figure 3 shows the same comparison while performing feature selection. These figures show the mean accuracy across all subjects while performing 4 different binary classifications based on pain levels. The final scores are presented in Table 6 and Table 7 below.

XSL•FO
RenderX

Figure 2. Validation accuracy of all classifiers on BioVid features. BL: baseline; PL: pain level; KNN: k-nearest neighbor; RF: random forest; SVM: support vector machine; XGB: XGBoost.

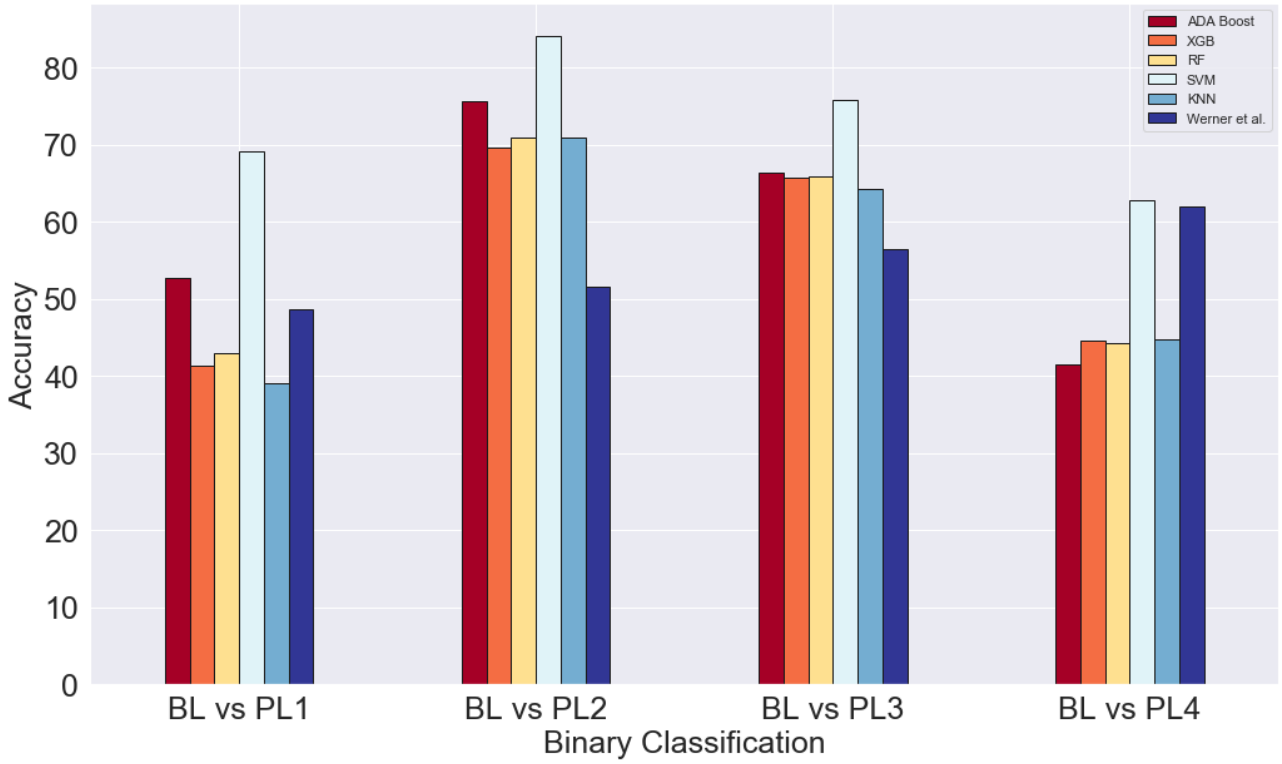
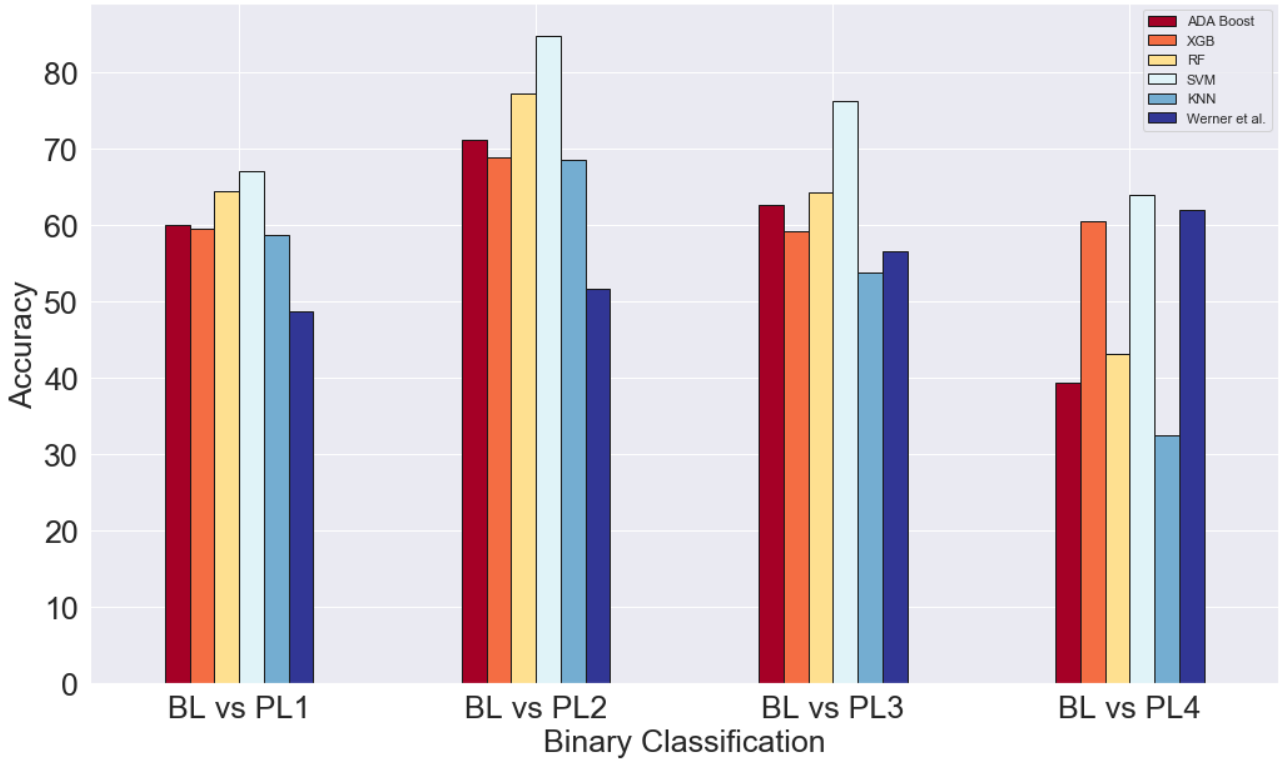


Figure 3. Validation accuracy of all classifiers on top 8 features. BL: baseline; PL: pain level; KNN: k-nearest neighbor; RF: random forest; SVM: support vector machine; XGB: XGBoost.



XSL•FO
RenderX

Table 6. Validation accuracy of BioVid features.

Binary classification	AdaBoost	XGBoost	RF[a]	SVM[b]	KNN[c]	Werner et al
BL[d] vs PL1[e]	52.63	41.35	42.97	69.16	39.06	48.7
BL vs PL2	75.68	69.57	70.84	84.14	70.92	51.6
BL vs PL3	66.33	65.73	65.94	75.73	64.20	56.5
BL vs PL4	41.53	44.55	44.24	62.72	44.68	62.0

[a]RF: random forest.

[b]SVM: support vector machine.

[c]KNN: k-nearest neighbor.

[d]BL: baseline.

[e]PL: pain level.

Table 7. Validation accuracy of top 8 features.

Binary classification	AdaBoost	XGBoost	RF[a]	SVM[b]	KNN[c]	Werner et al
BL[d] vs PL1[e]	59.94	59.46	64.37	67.03	58.61	48.7
BL vs PL2	71.06	68.85	77.19	84.79	68.54	51.6
BL vs PL3	62.63	59.22	64.29	76.18	53.76	56.5
BL vs PL4	39.29	60.44	43.17	63.86	32.51	62.0

[a]RF: random forest.

[b]SVM: support vector machine.

[c]KNN: k-nearest neighbor.

[d]BL: baseline.

[e]PL: pain level.

We were able to achieve the highest accuracy on the SVM classifier for both settings, with and without feature selection. Moreover, there is no noteworthy difference in the performance of the SVM classifier in both settings. However, while comparing the other classifiers, it is evident that there is a great improvement in performance while using feature selection in the BL versus PL1 category. The performances of the AdaBoost, XGBoost, RF, and KNN classifiers have a marked increase of about 12% on average when compared to their counterparts without feature selection. However, there is a slight decrease in the AdaBoost and RF classifiers and significant decrease in the KNN performance in the BL versus PL4 category. On the other hand, there is an improvement of about 16% in the XGBoost classifier while performing feature selection in the BL versus PL4 category. We speculate that the lower accuracy scores could be due to the relatively smaller number of training examples available from the downsampled PL4. On the flip side, due to the relative abundance of training examples from PL2, there is a spike in performance for all classifiers across both feature settings in the BL versus PL2 category.

While comparing our algorithm's performance to Werner et al [25], we can see that our SVM classifier fares significantly better than their model. The SVM classifier outperforms their model by an average of 20% across both feature settings for the first three pain categories (BL vs PL1, BL vs PL2, and BL vs PL3). Conversely, there is only a slight increase in performance across both feature settings in the BL versus PL4 category.

Discussion

Strengths

To the best of our knowledge, this is the first study that uses ECG signals from real postoperative adult patients for the purpose of developing an automatic pain assessment tool. Moreover, the use of weak supervision in our data labeling process is a novel approach that has not been implemented in pain assessment studies before. It eliminates the need for constantly asking patients for their pain levels and therefore reduces the burden placed on them during the trials. The accuracy scores for this data set with the focus of optimizing it for AUC using the SVM classifier, especially in the first three pain categories, are considerably higher than the scores achieved by Werner et al for both feature settings (with and without feature selection). We also achieve comparable results for the last pain category (BL vs PL4). Furthermore, the use of feature selection in our procedure helps determine the most informative features and reduces the complexity of our pain models. We were able to identify the 8 most informative features and improve the performance of our models in the process.

Limitations

The main limitation in our algorithm is the presence of noise, in the form of motion artifacts, in our physiological data. Since we collect data from real postoperative patients in a clinical setting, they were allowed to move more freely when compared to experiments performed in a laboratory setting on healthy subjects. The presence of noise diminished the quality of our

XSL•FO
RenderX

data. Thus, this negatively impacted the performance of our algorithm.

Another limitation in our experiments is the presence of imbalanced labels in each patient's data. Since we did not collect data in a laboratory setting, most patients did not report all the different pain levels during the trials. Most noticeably, this led to a relatively smaller number of labeled examples from the highest pain level (PL4). This consequently decreased the performance accuracy for that pain category (BL vs PL4). In a controlled laboratory setting, one can design the study to force the pain intensity levels to be balanced, which is not feasible in real settings.

Furthermore, we could not find a significant difference between different pain levels in our study. We believe this is due to the fact that variations in ECG signals in response to different pain levels are much harder to distinguish in comparison to different pain levels versus baseline. Moreover, it is worth mentioning that the state of the art in pain assessment focuses on comparing baseline with other pain levels (eg, Werner et al [25]). We believe the reason is to find out if the patient has pain (baseline vs other pain levels).

Conclusions

The experiments proposed in this study show the viability of our pain assessment algorithm on data from postoperative patients. The use of weak supervision for labeling and feature extraction improves the robustness of our approach. We plan to incorporate multimodal pain assessment methods to further improve our performance and robustness.

Acknowledgments

This research was supported by the Academy of Finland project, Personalized Pain Assessment System based on IoT (313488).

Conflicts of Interest

None declared.

References

1. Tompkins DA, Hobelmann JG, Compton P. Providing chronic pain management in the "Fifth Vital Sign" Era: Historical and treatment perspectives on a modern-day medical dilemma. Drug Alcohol Depend 2017 Dec 01;173 Suppl 1:S11-S21 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.12.002] [Medline: 28363315]
2. Aydede M. Defending the IASP Definition of Pain. The Monist 2017 Sep 01;100(4):439-464. [doi: 10.1093/monist/onx021]
3. Merskey H. Pain terms: a list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy. Pain 1979 Jun;6(3):249. [Medline: 460932]
4. Williams ACDC. Facial expression of pain: an evolutionary account. Behav Brain Sci 2002 Aug;25(4):439-55; discussion 455. [doi: 10.1017/s0140525x02000080] [Medline: 12879700]
5. Raffaeli W, Arnaudo E. Pain as a disease: an overview. J Pain Res 2017;10:2003-2008 [FREE Full text] [doi: 10.2147/JPR.S138864] [Medline: 28860855]
6. Breivik H, Borchgrevink P, Allen S, Rosseland L, Romundstad L, Hals EKB, et al. Assessment of pain. Br J Anaesth 2008 Jul;101(1):17-24 [FREE Full text] [doi: 10.1093/bja/aen103] [Medline: 18487245]
7. Stites M. Observational pain scales in critically ill adults. Crit Care Nurse 2013 Jun;33(3):68-78. [doi: 10.4037/ccn2013804] [Medline: 23727853]
8. Barr J, Fraser GL, Puntillo K, Ely EW, Gélinas C, Dasta JF, American College of Critical Care Medicine. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306. [doi: 10.1097/CCM.0b013e3182783b72] [Medline: 23269131]
9. Herr K, Coyne PJ, McCaffery M, Manworren R, Merkel S. Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations. Pain Manag Nurs 2011 Dec;12(4):230-250. [doi: 10.1016/j.pmn.2011.10.002] [Medline: 22117755]
10. Kasaeyan Naeini E, Shahhosseini S, Subramanian A, Yin T, Rahmani A, Dutt N. An Edge-Assisted and Smart System for Real-Time Pain Monitoring. 2019 Nov 25 Presented at: IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2019; Arlington, VA, USA p. 47.
11. Ben-Israel N, Kliger M, Zuckerman G, Katz Y, Edry R. Monitoring the nociception level: a multi-parameter approach. J Clin Monit Comput 2013 Dec;27(6):659-668. [doi: 10.1007/s10877-013-9487-9] [Medline: 23835792]
12. Gruss S, Treister R, Werner P, Traue HC, Crawcour S, Andrade A, et al. Pain Intensity Recognition Rates via Biopotential Feature Patterns with Support Vector Machines. PLoS One 2015;10(10):e0140330 [FREE Full text] [doi: 10.1371/journal.pone.0140330] [Medline: 26474183]
13. Treister R, Kliger M, Zuckerman G, Goor Aryeh I, Eisenberg E. Differentiating between heat pain intensities: the combined effect of multiple autonomic parameters. Pain 2012 Sep;153(9):1807-1814. [doi: 10.1016/j.pain.2012.04.008] [Medline: 22647429]
14. Vatankhah M, Toliyat A. Pain Level Measurement Using Discrete Wavelet Transform. IJET 2016 May;8(5):380-384. [doi: 10.7763/IJET.2016.V8.917]

XSL•FO
RenderX

15. Kasaeyan Naeini E, Jiang M, Syrjälä E, Calderon M, Mieronkoski R, Zheng K, et al. Prospective Study Evaluating a Pain Assessment Tool in a Postoperative Environment: Protocol for Algorithm Testing and Enhancement. JMIR Res Protoc 2020 Jul 01;9(7):e17783 [FREE Full text] [doi: 10.2196/17783] [Medline: 32609091]
16. Appelhans BM, Luecken LJ. Heart rate variability and pain: associations of two interrelated homeostatic processes. Biol Psychol 2008 Feb;77(2):174-182. [doi: 10.1016/j.biopsycho.2007.10.004] [Medline: 18023960]
17. Jiang M, Mieronkoski R, Rahmani A, Hagelberg N, Salanterä S, Liljeberg P. Ultra-short-term analysis of heart rate variability for real-time acute pain monitoring with wearable electronics. 2017 Dec 18 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2017; Kansas City, MO, USA p. 1025.
18. Lopez-Martinez D, Picard R. Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:5624-5627. [Medline: 30441611]
19. Koenig J, Jarczok MN, Ellis RJ, Hillecke TK, Thayer JF. Heart rate variability and experimentally induced pain in healthy adults: a systematic review. Eur J Pain 2014 Mar;18(3):301-314. [doi: 10.1002/j.1532-2149.2013.00379.x] [Medline: 23922336]
20. Sarker V, Jiang M, Gia T, Anzanpour A, Rahmani A, Liljeberg P. Portable multipurpose bio-signal acquisition and wireless streaming device for wearables. 2017 Apr 12 Presented at: IEEE Sensors Applications Symposium (SAS); 2017; Glassboro, NJ, USA p. 2017-2017. [doi: 10.1109/sas.2017.7894053]
21. Jiang M, Mieronkoski R, Syrjälä E, Anzanpour A, Terävä V, Rahmani AM, et al. Acute pain intensity monitoring with the classification of multiple physiological parameters. J Clin Monit Comput 2019 Jun;33(3):493-507 [FREE Full text] [doi: 10.1007/s10877-018-0174-8] [Medline: 29946994]
22. Ratner A, De Sa C, Wu S, Selsam D, Ré C. Data Programming: Creating Large Training Sets, Quickly. Adv Neural Inf Process Syst 2016 Dec;29:3567-3575 [FREE Full text] [Medline: 29872252]
23. Ratner A, Bach SH, Ehrenberg H, Fries J, Wu S, Ré C. Snorkel: Rapid Training Data Creation with Weak Supervision. Proceedings VLDB Endowment 2017 Nov;11(3):269-282 [FREE Full text] [doi: 10.14778/3157794.3157797] [Medline: 29770249]
24. Sesay M, Robin G, Tauzin-Fin P, Sacko O, Gimbert E, Vignes JR, et al. Responses of heart rate variability to acute pain after minor spinal surgery: optimal thresholds and correlation with the numeric rating scale. J Neurosurg Anesthesiol 2015 Apr;27(2):148-154. [doi: 10.1097/ANA.0000000000000102] [Medline: 25105826]
25. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue H. Automatic Pain Recognition from Video and Biomedical Signals. 2014 Dec Presented at: 22nd International Conference on Pattern Recognition; 2014; Stockholm, Sweden p. 4582.
26. Johnson C. Measuring Pain. Visual Analog Scale Versus Numeric Pain Scale: What is the Difference? J Chiropr Med 2005;4(1):43-44 [FREE Full text] [doi: 10.1016/S0899-3467(07)60112-8] [Medline: 19674646]
27. Laitala J, Jiang M, Syrjälä E, Kasaeyan Naeini E, Airola A, Rahmani A, et al. Robust ECG R-Peak Detection Using LSTM. 2020 Mar Presented at: Proceedings of the 35th Annual ACM Symposium on Applied Computing; 2020; Brno Czech Republic p. 1104-1111.
28. Gomes P, Margaritoff P, Plácido DSH. pyHRV: Development and Evaluation of an Open-Source Python Toolbox for Heart Rate Variability (HRV). In: Proc. Int'l Conf. on Electrical, Electronic and Computing Engineering (IcETRAN). 2019 Presented at: International Conference on Electrical, Electronic and Computing Engineering (IcETRAN); 2019; Serbia p. 822-828.
29. Tang J, Alelyani S, Liu H. Feature Selection for Classification: A Review. In: Data Classification: Algorithms and Applications. Oxfordshire: Taylor and Francis; 2014.

Abbreviations

APS: Acute Pain Service
AUC: area under the curve
BL: baseline
ECG: electrocardiography
FFT: fast Fourier transform
HF: high-frequency
HRV: heart rate variability
IoT: Internet of Things
KNN: k-nearest neighbor
LF: low-frequency
LOOCV: leave-one-out cross-validation
NaN: not a number
NRS: Numerical Rating Scale
PL: pain level
RF: random forest
RMSSD: root mean square of successive differences

XSL•FO
RenderX

SVM: support vector machine
TENS: transcutaneous electrical nerve stimulation
UCI: University of California, Irvine
VAS: Visual Analogue Scale
VLF: very low-frequency

Edited by R Kukafka; submitted 19.10.20; peer-reviewed by D Firl, D Schwarz, P Kocbek; comments to author 23.02.21; revised version received 06.04.21; accepted 25.04.21; published 28.05.21.

Please cite as:
Kasaeyan Naeini E, Subramanian A, Calderon MD, Zheng K, Dutt N, Liljeberg P, Salantera S, Nelson AM, Rahmani AM
Pain Recognition With Electrocardiographic Features in Postoperative Patients: Method Validation Study
J Med Internet Res 2021;23(5):e25079
URL: https://www.jmir.org/2021/5/e25079
doi:10.2196/25079
PMID:34047710

©Emad Kasaeyan Naeini, Ajan Subramanian, Michael-David Calderon, Kai Zheng, Nikil Dutt, Pasi Liljeberg, Sanna Salantera, Ariana M Nelson, Amir M Rahmani. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.05.2021. 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